

A NEW PRIOR FOR DISCRETE DAG MODELS WITH A RESTRICTED SET OF DIRECTIONS

HÉLÈNE MASSAM AND JACEK WESOŁOWSKI

ABSTRACT. In this paper, we first develop a new family of conjugate prior distributions for the cell parameters of discrete graphical models Markov with respect to a set \mathcal{P} of moral directed acyclic graphs with skeleton a given decomposable graph G . Such families arise when the set of conditional independences between discrete variables is given and can be represented by a decomposable graph and additionally, the direction of certain edges is imposed by the practitioner. This family, which we call the \mathcal{P} -Dirichlet, is a generalization of the hyper Dirichlet given in Dawid and Lauritzen (1993): it keeps the strong directed hyper Markov property for every DAG in \mathcal{P} but increases the flexibility in the choice of its parameters, i.e. the hyper parameters.

Our second contribution is a characterization of the \mathcal{P} -Dirichlet, which yields, as a corollary, a characterization of the hyper Dirichlet and a characterization of the Dirichlet also. Like that given by Geiger and Heckerman (1997), our characterization of the Dirichlet is based on local and global independence of the probability parameters but we need not make the assumption of the existence of a positive density function. We use the method of moments for our proofs.

1. INTRODUCTION

The Dirichlet distribution and distributions derived from the Dirichlet are essential ingredients of Bayesian inference in the analysis of discrete data. For high-dimensional data, Dirichlet-type distributions are often used in conjunction with graphical models. Let $V = \{1, \dots, d\}$ be a finite set indexing the variables. A graphical model for the discrete random variable $X = (X_v, v \in V)$ is a statistical model where the dependences between $X_v, v \in V$ are expressed by means of a graph G . We will assume here that the data is gathered under the form of a d -dimensional contingency table and that the cell counts follow a multinomial distribution with cell probabilities $(p(\underline{i}), \underline{i} \in \mathcal{I})$ where \mathcal{I} is the set of cells in the contingency table.

If there are no independences between the variables $X_v, v \in V$, then the cell counts follow a standard multinomial distribution and the Diaconis-Ylvisaker (1979) prior distribution on $(p(\underline{i}), \underline{i} \in \mathcal{I})$ is the Dirichlet distribution. If the conditional independences between the variables can be expressed by means of a directed acyclic graph (henceforth abbreviated DAG), the usual priors are independent Dirichlet prior distributions on

$$(1) \quad \left(p(i_v | X_{\mathbf{p}_v} = \underline{i}_{\mathbf{p}_v}), i_v \in \mathcal{I}_v \right), \underline{i}_{\mathbf{p}_v} \in \mathcal{I}_{\mathbf{p}_v}, v \in V,$$

1991 *Mathematics Subject Classification.* 62H17, 62F15, 62E99.

Key words and phrases. Bayesian learning, strong directed hyper Markov, conjugate priors, prior knowledge, hyper Dirichlet distribution.

where, for $A \subset V$, \mathcal{I}_A is the finite set of values that $X_A = (X_v, v \in A)$ can take, \mathbf{p}_v denotes the set of parents of vertex v in the DAG and $\underline{i}_{\mathbf{p}_v} = (i_l, l \in \mathbf{p}_v)$ denotes the \mathbf{p}_v -marginal cell. This means that the probability parameters associated with each node are mutually independent (global independence) and for a given node, the parameters associated with various instances of its parents are also mutually independent (local parameter independence).

Geiger and Heckerman (1997, Theorem 3) have shown that if the distribution of positive random variables that sum to 1 has a strictly positive density and possesses the property of local and global independence for the two directions $\{1, 2, \dots, d-1, d\}$ and $\{d, 1, 2, \dots, d-1\}$ of a complete DAG on V , this distribution must be the Dirichlet distribution. This is a characterization of the Dirichlet through local and global independence using two particular complete DAG's. This result has important practical ramifications for the choice of the hyper parameters of the Dirichlet priors for model selection, as emphasized by Geiger and Heckerman (1997). Indeed, for model selection in the class of models Markov with respect to DAG's on V , if we request that a node with the same parents in two distinct DAG's has identical prior distribution on the parameters associated with this node in both structures (this is called parameter modularity) and if we also request local and global independence of the parameters for all possible DAG's, then the priors on the variables in (1) must all be derived from one single Dirichlet distribution $\text{Dir}(\alpha(i), i \in I)$ on $(p(\underline{i}), \underline{i} \in \mathcal{I})$. The hyper parameters $\alpha(i)$ of such a Dirichlet are of the form

$$(2) \quad \alpha(i) = \alpha \theta(i)$$

where $\theta(i)$ is the expected value of $p(i)$ and α , called the *equivalent sample size*, (which we can think of as the total cell count in a fictive contingency table representing prior knowledge) will be added to the actual total cell count $N = \sum_{\underline{i} \in \mathcal{I}} n(\underline{i})$ in posterior inference. This α will therefore represent our confidence in our choice of the particular fictive table. Most of the time, in practice and in the absence of prior expert information, we take $\theta(i) = \frac{1}{|\mathcal{I}|}$ where $|\mathcal{I}|$ is the total number of cells. This implies that, often in practice, our flexibility in the choice of the hyper parameters is restricted to the choice of α .

If there are conditional independences between the variables and they can be expressed by means of a decomposable undirected graph G , then the multinomial distribution on the cell counts is Markov with respect to G and the Diaconis-Ylvisaker (1979) prior on $(p(\underline{i}), \underline{i} \in \mathcal{I})$ is the hyper Dirichlet defined by David and Lauritzen (1993). A simple calculation (see equations (14) in Section 2) shows that the parameters of the hyper Dirichlet have an interpretation of the type (2) and we are therefore faced with the same lack of flexibility in the choice of the hyper parameters: again, we are restricted to the choice of one parameter only. Moreover it is intuitively clear (and it can be proved, see Theorem 4.6) that if, for any DAG with skeleton G , Markov equivalent to G , we make the change of variable from $p(i), i \in \mathcal{I}$ to the variables in equation (1), i.e., $(p(i_v | X_{\mathbf{p}_v} = \underline{i}_{\mathbf{p}_v}), i_v \in \mathcal{I}_v), \underline{i}_{\mathbf{p}_v} \in \mathcal{I}_{\mathbf{p}_v}, v \in V$, then, for each of these DAGs, the distribution induced from the hyper Dirichlet also possesses the property of local and global independence. One can then be led to think that the property of local and global independence in all possible directions imposes strict restrictions on the choice of the hyper parameters of the prior on $p(i), i \in \mathcal{I}$.

In this paper, we therefore consider a family \mathcal{P} of DAG's with skeleton a decomposable graph G and Markov equivalent to G but we do not include in this family all the DAG's Markov equivalent to G : \mathcal{P} is a restricted family of such DAG's. For the set of discrete multinomial models Markov with respect to any of the DAG's in \mathcal{P} , we build a new family of conjugate prior distributions for the parameters $(\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ of this set of models. As we show in Section 5.3, such a family of priors will be the right family of prior distributions to use if we are given by an expert the set of conditional independences (those given by a decomposable graph G) and the restriction that some edges must have a given direction. In that case, there is no need to require local and global independence for all DAG's Markov with respect to G . We only need to require it for those DAG's containing the compulsory edges. This new family of priors possesses the property of local and global independence and also, as we shall see later, the strong directed hyper Markov property. Using these priors thus facilitates posterior inference. Moreover, because of the restriction on \mathcal{P} , it offers more hyper parameter flexibility than the hyper Dirichlet. The development and study of this new family of priors, called \mathcal{P} -Dirichlet, is the first contribution of this paper. The second contribution is a characterization (Theorem 6.1) of this new family through local and global independence. As corollaries, we obtain a characterization of the hyper-Dirichlet and also the characterization of the Dirichlet as given by Geiger and Heckerman but without assuming the existence of a density for $\mathbf{p} = (\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$.

The remainder of this paper is organized as follows. The \mathcal{P} -Dirichlet family of priors on $(p(\underline{i}), \underline{i} \in \mathcal{I})$, is defined in Section 3, through local and global independence with respect to the collection \mathcal{P} of DAGs. In Section 4, we first give an example of a \mathcal{P} -Dirichlet family and then we derive a general formula for the moments of the \mathcal{P} -Dirichlet which will be used in its characterization, Theorem 6.1. In the expression of the moments, we will see that the role played by the collection of cliques and separators in the hyper Dirichlet is played by two larger collections, \mathfrak{Q} and \mathfrak{P} , of complete subsets of the decomposable graph G which include respectively, the set of cliques and separators. These additional complete subsets yield more parameters for the \mathcal{P} -Dirichlet distribution and thus increase flexibility. In Section 5, we give the dimension of the \mathcal{P} -Dirichlet family and show that it is always larger than (or equal to) the dimension of the hyper Dirichlet family with the same skeleton. We also show that the \mathcal{P} -Dirichlet is a conjugate family for the multinomial distribution Markov with respect to G and that it has the strong directed hyper Markov for any DAG in \mathcal{P} . In Section 6, we show that, when \mathcal{P} is what we call a *separating* family of DAG's, local and global independence characterizes the \mathcal{P} -Dirichlet distribution. We will look at special cases. When G is a tree, for the \mathcal{P} -Dirichlet to be the hyper Dirichlet, \mathcal{P} needs only be the set of DAG's with their root at a leaf of the tree, not the set of all possible DAG's Markov equivalent to G . When G is complete, a separating family of DAG's can be formed by taking the orders $\{1, 2, \dots, d\}$ and $\{d, 1, 2, \dots, d-1\}$ as was done by Geiger and Heckerman (1997) but this is not the only separating family and the characterization of the Dirichlet can be obtained through any separating family of DAG's on a complete graph G .

We conclude with Section 7 summarizing the contributions of this paper. Proofs of the moment formula in Theorem 4.3 and the characterization, Theorem 6.1 are given in the Appendix, Section 8. The next section is devoted to preliminaries.

2. PRELIMINARIES

2.1. Graph theoretical notions. Let $V = \{1, \dots, d\}$ be a finite set of indices for the d criteria defining the contingency table with our data. We assume that the criterion labelled by $v \in V$ can take values in a finite set I_v . Let

$$\mathcal{I} = \prod_{v \in V} \mathcal{I}_v$$

be the set of cells $\underline{i} = (i_v, v \in V)$. If $D \subset V$ and $\underline{i} \in \mathcal{I}$ we write $\underline{i}_D = (i_v, v \in D)$ for the D -marginal cell. Let $G = (V, E)$ be an undirected graph as defined above. A graph is said to be decomposable if it does not have any cycle of length greater than or equal to 4. A subset D of V is said to be complete if all vertices of D are linked to each other with an edge. A clique is a complete subset maximal with respect to inclusion. Let \mathcal{C} denote the set of cliques. For any given order C_1, \dots, C_K of the cliques of G we will define

$$H_j = \cup_{l=1}^j C_l, \quad S_j = C_j \cap H_{j-1} \quad \text{and} \quad R_j = C_j \setminus H_{j-1} = C_j \setminus S_j, \quad j = 1, \dots, K,$$

with $S_1 = \emptyset$, called respectively the j -th history, j -th separator and j -th residual. An order C_1, \dots, C_K of the cliques of G is said to be perfect if for any $j > 1$, there exists an $i < j$ such that

$$(\cup_{l=1}^{j-1} C_l) \cap C_j \subset C_i.$$

It can be shown (Proposition 2.17, Lauritzen, 1996) that a graph G is decomposable if and only if its set of cliques admits a perfect order. It can also be shown that the set of separators \mathcal{S} associated with a perfect ordering of the cliques is independent of the perfect ordering considered and it is the set of minimal separators for the graph G .

For a given decomposable G , we will consider the DAG's with skeleton G which are moral. A DAG is said to be moral if when v and v' are both parents of w , then there is an edge between v and v' . In this paper, we will deal exclusively with moral DAG's. Note that any DAG with skeleton G is in one-to-one correspondence with the function $\mathbf{p} : V \rightarrow 2^V$ which describes the parents of each vertex or with the function $\mathbf{c} : V \rightarrow 2^V$ which describes the children of each vertex. We will write \mathbf{p}_v and \mathbf{c}_v for the set of parents and children of $v \in V$, respectively. The set of descendants, that is children of children and so on, of a vertex v is denoted \mathbf{d}_v and the set of nondescendants will be denoted \mathbf{nd}_v . Additionally denote $\mathbf{q}_v = \mathbf{p}_v \cup \{v\}$ for any $v \in V$ and let $\mathbf{q} : V \rightarrow 2^V$ denote the corresponding mapping. Note that for any $v \in S$ for some separator S , due to morality, there exists a unique clique $C \supset S$ such that $\mathbf{p}_v \subset C$.

We will now define some additional graph-theoretical notions needed in this work.

Definition 2.1. For $S \in \mathcal{S}$ and $C \in \mathcal{C}$, we say that C and S are paired by a given perfect ordering $o = (C_1, \dots, C_K)$ of the cliques if

$$\exists l \in \{1, \dots, K\} : \quad S = S_l, \quad C = C_l$$

(here we allow that $S = S_1 = \emptyset$). If S and C are paired by o , we write $S \xrightarrow{o} C$.

Definition 2.2. Given a parent function \mathbf{p} , a perfect order $o = (C_1, \dots, C_K)$ of the cliques is said to be \mathbf{p} -perfect if for any $l = 1, \dots, K$ there exists a $v \in C_l \setminus S_l$ such that $S_l = \mathbf{p}_v$.

With a given $\mathbf{p} \in \mathcal{P}$ and a \mathbf{p} -perfect order $o = (C_1, \dots, C_K)$ of the cliques, we associate the following numbering of the vertices: $\forall l \in \{1, \dots, K\}$

$$(3) \quad v = \begin{cases} v_{l, s_l+1} & \text{iff } \mathbf{p}_v = S_l, \\ v_{l, j} & \text{iff } \mathbf{p}_v = \mathbf{q}_{v_{l, j-1}}, j = s_l + 2, \dots, c_l, \end{cases}$$

where c_l and s_l denote the cardinality of C_l and S_l respectively. We note that if $v = v_{l, c_l}$ then $\mathbf{q}_v = C_l$, $l = 1, \dots, K$ and if $v = v_{1, 1}$, the source vertex of \mathbf{p} , then $\mathbf{p}_v = \emptyset$.

2.2. Markov properties and the hyper Dirichlet. Let $\mathbf{X} = (X_1, \dots, X_d)$ be a discrete random vector with variables X_v taking values in \mathcal{I}_v , $v \in V$. Given an undirected graph G , the distribution of X is said to be Markov with respect to the G if X_i and X_j are independent given $\mathbf{X}_{V \setminus \{i, j\}}$ whenever the pair $\{i, j\}$ does not belong to E .

Given a DAG \mathcal{G} , we say that \mathbf{X} is Markov with respect to \mathcal{G} if, for any $v \in V$,

$$(4) \quad X_v \perp \mathbf{X}_{\text{nd}_v} \mid \mathbf{X}_{\mathbf{p}_v}.$$

Therefore if \mathbf{X} is Markov with respect to \mathcal{G} , we have

$$(5) \quad p(\underline{i}) := \mathbb{P}(\mathbf{X} = \underline{i}) = \mathbb{P}(X_v = i_v, v \in V) = \prod_{v \in V} \mathbb{P}(X_v = i_v \mid \mathbf{X}_{\mathbf{p}_v} = \underline{i}_{\mathbf{p}_v}).$$

The two formulations (4) and (5) are equivalent.

In this paper, we consider only DAGs \mathcal{G} with a decomposable skeleton G and which encode the same conditional independences as G . By Lemma 3.21 of Lauritzen (1996), such a DAG is necessarily moral and it has been shown in Andersson, Madigan, Perlman and Triggs (1997) that, for this DAG, we can use a numbering of the vertices of the type (3). With this numbering the two models, Markov with respect to G and \mathcal{G} , define the same set of conditional independences: they are said to be Markov equivalent. It is well known that for any $\underline{i} \in \mathcal{I}$, we have the following decomposition of the probability that $\mathbf{X} = \underline{i} = (i_1, \dots, i_d)$,

$$(6) \quad \mathbb{P}(\mathbf{X} = \underline{i}) = \mathbb{P}(X_v = i_v, v \in V) = \frac{\prod_{l=1}^K \mathbb{P}(\mathbf{X}_{C_l} = \underline{i}_{C_l})}{\prod_{l=2}^K \mathbb{P}(\mathbf{X}_{S_l} = \underline{i}_{S_l})}.$$

Then, if the cell counts $(N(\underline{i}), \underline{i} \in \mathcal{I})$ follow a Multinomial $(N, p(\underline{i}), \underline{i} \in \mathcal{I})$ with $p(\underline{i})$ as given in (5), the density of $(N(\underline{i}), \underline{i} \in \mathcal{I})$ with respect to the counting measure is proportional to

$$(7) \quad \begin{aligned} \prod_{\underline{i} \in \mathcal{I}} p(\underline{i})^{n(\underline{i})} &= \frac{\prod_{l=1}^K \prod_{\underline{i}_{C_l} \in \mathcal{I}_{C_l}} \mathbb{P}(\mathbf{X}_{C_l} = \underline{i}_{C_l})^{n(\underline{i}_{C_l})}}{\prod_{l=2}^K \prod_{\underline{i}_{S_l} \in \mathcal{I}_{S_l}} \mathbb{P}(\mathbf{X}_{S_l} = \underline{i}_{S_l})^{n(\underline{i}_{S_l})}} \\ &= \prod_{v \in V} \prod_{\mathbf{p}_v \in \mathcal{I}_{\mathbf{p}_v}} \prod_{i_v \in \mathcal{I}_v} \mathbb{P}(\mathbf{X}_v = i_v \mid \mathbf{X}_{\mathbf{p}_v} = \underline{i}_{\mathbf{p}_v})^{n(\underline{i}_{\mathbf{p}_v})} \end{aligned}$$

where the two equalities follow from (6) and (5), respectively.

We note that for any $l = 1, \dots, K$ the marginal counts satisfy

$$(8) \quad \sum_{\underline{i}_{C_l} \in \mathcal{I}_{C_l}} n(\underline{i}_{C_l}) = N, \quad \sum_{\underline{j}_{C_l} \in \mathcal{I}_{C_l} : \underline{j}_{S_l} = \underline{i}_{S_l}} n(\underline{j}_{C_l}) = n(\underline{i}_{S_l}), \quad \sum_{\underline{i}_{S_l} \in \mathcal{I}_{S_l}} n(\underline{i}_{S_l}) = N$$

and, for $v \in C_l$ and $\mathbf{q}_v \subseteq C_l$,

$$(9) \quad \sum_{\underline{j}_{C_l} \in I_{C_l} : \underline{j}_{\mathbf{q}_v} = \underline{i}_{\mathbf{q}_v}} n(\underline{j}_{C_l}) = n(\underline{i}_{\mathbf{q}_v}), \quad \underline{i}_{\mathbf{q}_v} \in \mathcal{I}_{\mathbf{q}_v}.$$

In Bayesian inference we view the parameter of the multinomial distribution as a random vector. Let $\mathbf{p} = (\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ be the vector of random cell probabilities, that is $\mathbf{p}(\underline{i}) > 0$ for any $\underline{i} \in \mathcal{I}$ and $\sum_{\underline{i} \in \mathcal{I}} \mathbf{p}(\underline{i}) = 1$. Note that we write $\mathbf{p} = (\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ for the random vector of cell probabilities, and $p = (p(\underline{i}), \underline{i} \in \mathcal{I})$ for its value. Let

$$\mathbb{P}_p(\mathbf{X} = \underline{i}) := \mathbb{P}(\mathbf{X} = \underline{i} | \mathbf{p} = p) = p(\underline{i}), \quad \underline{i} \in \mathcal{I}.$$

The random variables $\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i}) = \mathbb{P}(\mathbf{X} = \underline{i} | \mathbf{p})$, $\underline{i} \in \mathcal{I}$, are the variables of interest in this paper.

As mentioned in the introduction the Diaconis-Ylvisaker conjugate prior on \mathbf{p} is the hyper Dirichlet first identified by Dawid and Lauritzen (1993) with density

$$(10) \quad \frac{\Gamma(\alpha) \prod_{l=2}^K \prod_{\underline{i}_{S_l} \in \mathcal{I}_{S_l}} \Gamma(\alpha_{\underline{i}_{S_l}}^{S_l})}{\prod_{l=1}^K \prod_{\underline{i}_{C_l} \in \mathcal{I}_{C_l}} \Gamma(\alpha_{\underline{i}_{C_l}}^{C_l})} \frac{\prod_{l=1}^K \prod_{\underline{i}_{C_l} \in \mathcal{I}_{C_l}} (p_{\underline{i}_{C_l}}^{C_l})^{\alpha_{\underline{i}_{C_l}}^{C_l} - 1}}{\prod_{l=2}^K \prod_{\underline{i}_{S_l} \in \mathcal{I}_{S_l}} (p_{\underline{i}_{S_l}}^{S_l})^{\alpha_{\underline{i}_{S_l}}^{S_l} - 1}},$$

where $p_{\underline{i}_{C_l}}^{C_l}$ and $p_{\underline{i}_{S_l}}^{S_l}$ are the values of the random variables $\mathbf{p}_{\underline{i}_{C_l}}^{C_l} = \mathbb{P}_{\mathbf{p}}(\mathbf{X}_{C_l} = \underline{i}_{C_l})$ and $\mathbf{p}_{\underline{i}_{S_l}}^{S_l} = \mathbb{P}_{\mathbf{p}}(\mathbf{X}_{S_l} = \underline{i}_{S_l})$, $l = 1, \dots, K$, respectively, and where the hyperparameters satisfy constraints parallel to (8), namely

$$(11) \quad \sum_{\underline{i}_{C_l} \in \mathcal{I}_{C_l}} \alpha_{\underline{i}_{C_l}}^{C_l} = \alpha, \quad \sum_{\underline{j}_{C_l} \in \mathcal{I}_{C_l} : \underline{j}_{S_l} = \underline{i}_{S_l}} \alpha_{\underline{j}_{C_l}}^{C_l} = \alpha_{\underline{i}_{S_l}}^{S_l}, \quad \sum_{\underline{i}_{S_l} \in \mathcal{I}_{S_l}} \alpha_{\underline{i}_{S_l}}^{S_l} = \alpha.$$

Unlike the case of the Dirichlet on the complete graph, this density is typically defined on a complicated manifold determined not only by summation to one but also by all the conditional independence properties encoded in G . This is one of the reasons our approach is through moments and not densities.

We now want to recall the expression of the moments of the hyper Dirichlet. For simplicity, let us write

$$\alpha(G, \mathcal{I}) = (\alpha, \alpha_{\underline{k}}^{C_l}, \underline{k} \in \mathcal{I}_{C_l}, l = 1, \dots, K, \alpha_{\underline{k}}^{S_l}, \underline{k} \in \mathcal{I}_{S_l}, l = 2, \dots, K).$$

For any d -way table $\underline{r} = (r_{\underline{i}}, \underline{i} \in \mathcal{I})$ of non-negative integers, we write the E -marginal counts as

$$r_{\underline{e}}^E = \sum_{\underline{i} : \underline{i}_E = \underline{e}} r(\underline{i}).$$

In particular, we will use $\underline{r}_{\underline{i}_{C_l}}^{C_l}$, $\underline{r}_{\underline{i}_{S_l}}^{S_l}$. By analogy with $\alpha(G, \mathcal{I})$, we will also use the notation $r(G, \mathcal{I})$ and write the normalizing constant of the hyper Dirichlet as

$$(12) \quad \mathcal{Z}(\alpha(G, \mathcal{I})) = \frac{\prod_{l=1}^K \prod_{\underline{i}_{C_l} \in \mathcal{I}_{C_l}} \Gamma(\alpha_{\underline{i}_{C_l}}^{C_l})}{\Gamma(\alpha) \prod_{l=2}^K \prod_{\underline{i}_{S_l} \in \mathcal{I}_{S_l}} \Gamma(\alpha_{\underline{i}_{S_l}}^{S_l})}$$

The moments are then equal to

$$(13) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i})]^{r_{\underline{i}}} = \mathbb{E} \frac{\prod_{l=1}^K \prod_{\underline{i}_{C_l} \in \mathcal{I}_{C_l}} \mathbb{P}_{\mathbf{p}}(\mathbf{X}_{C_l} = \underline{i}_{C_l})^{r_{\underline{i}_{C_l}}}}{\prod_{l=2}^K \prod_{\underline{i}_{S_l} \in \mathcal{I}_{S_l}} \mathbb{P}_{\mathbf{p}}(\mathbf{X}_{S_l} = \underline{i}_{S_l})^{r_{\underline{i}_{S_l}}}} = \frac{\mathcal{Z}(\alpha(G, \mathcal{I}) + r(G, \mathcal{I}))}{\mathcal{Z}(\alpha(G, \mathcal{I}))}.$$

In particular,

$$\mathbb{E} [\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i})] = \frac{\prod_{l=1}^K \Gamma(\alpha_{\underline{i}_{C_l}}^{C_l} + 1)}{\Gamma(\alpha + 1) \prod_{l=2}^K \Gamma(\alpha_{\underline{i}_{S_l}}^{S_l} + 1)} \frac{\Gamma(\alpha) \prod_{l=2}^K \Gamma(\alpha_{\underline{i}_{S_l}}^{S_l})}{\prod_{l=1}^K \Gamma(\alpha_{\underline{i}_{C_l}}^{C_l})} = \frac{\prod_{l=1}^K \alpha_{\underline{i}_{C_l}}^{C_l}}{\alpha \prod_{l=2}^K \alpha_{\underline{i}_{S_l}}^{S_l}}$$

Together with the constraints (11), this shows that we can write

$$(14) \quad \alpha_{\underline{i}_{C_l}}^{C_l} = \alpha \theta(i_C), \quad \alpha_{\underline{i}_{S_l}}^{S_l} = \alpha \theta(i_S),$$

where $\theta(i_C) = \mathbb{E} [p(i_C)]$ and $\theta(i_S) = \mathbb{E} [p(i_S)]$, a relationship similar to (2) as mentioned in the introduction.

3. \mathbf{p} -DIRICHLET AND \mathcal{P} -DIRICHLET DISTRIBUTIONS

Let $G = (V, E)$ be a decomposable graph. Let $\mathbf{X} = (X_1, \dots, X_d)$, $d \geq 2$, be a random vector which is Markov with respect to G and assumes values in the set $\mathcal{I} = \mathcal{I}_1 \times \dots \times \mathcal{I}_d$, where \mathcal{I}_j is a discrete set, with distinguished zero state \emptyset_j , $j = 1, 2, \dots, d$.

Since a moral DAG Markov equivalent to G is in 1-1 correspondence with the parent function $\mathbf{p} : V \rightarrow 2^V$ which describes the parents of each vertex $v \in V$, following (5), we have that

$$\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i}) = \prod_{v \in V} \mathbb{P}_{\mathbf{p}}(X_v = i_v | \mathbf{X}_{\mathbf{p}_v} = i_{\mathbf{p}_v}).$$

Letting $\mathbf{q}_v = \mathbf{p}_v \cup \{v\}$, $v \in V$, we can thus define the random variables

$$\mathbf{p}_{m|\underline{k}}^{v|\mathbf{p}_v} := \mathbb{P}_{\mathbf{p}}(X_v = m | \mathbf{X}_{\mathbf{p}_v} = \underline{k}) = \frac{\mathbf{p}^{\mathbf{q}_v}((k, m))}{\mathbf{p}^{\mathbf{p}_v}(\underline{k})}, \quad m \in \mathcal{I}_v, \underline{k} \in \mathcal{I}_{\mathbf{p}_v}, \quad v \in V,$$

where $\mathbf{p}^D(\underline{n}) = \sum_{\underline{j} \in \mathcal{I}: \underline{j}_D = \underline{n}} \mathbf{p}(\underline{j})$ for $\underline{n} \in \mathcal{I}_D$, $D \subset V$ and we have

$$(15) \quad \mathbf{p}(\underline{i}) = \prod_{v \in V} \mathbf{p}_{i_v|\underline{i}_{\mathbf{p}_v}}^{v|\mathbf{p}_v}, \quad \underline{i} \in \mathcal{I}.$$

We will say that a random vector $(\mathbf{p}(\underline{i}))_{\underline{i} \in \mathcal{I}}$ is associated with the graph G if it factorizes with respect to G , i.e. $p(i)$ can be written as

$$p(i) = \frac{\prod_{C \in \mathcal{C}} p(i_C)}{\prod_{S \in \mathcal{S}} p(i_S)}, \quad i \in \mathcal{I},$$

that is, if $\mathbf{X} = (X_1, \dots, X_d)$ is Markov with respect to G .

Definition 3.1. *The random vector $(\mathbf{p}(\underline{i}))_{\underline{i} \in \mathcal{I}}$ associated with the graph G has a \mathbf{p} -Dirichlet distribution if the random vectors $(\mathbf{p}_{m|\underline{k}}^{v|\mathbf{p}_v})_{m \in \mathcal{I}_v, \underline{k} \in \mathcal{I}_{\mathbf{p}_v}, v \in V}$, in representation (15) are independent and follow (classical) Dirichlet distributions.*

Recall that if a random vector $(\mathbf{p}_{m|\underline{k}}^{v|\mathbf{p}_v})_{m \in \mathcal{I}_v}$ has a classical Dirichlet distribution $\text{Dir}(\alpha_{m|\underline{k}}^{v|\mathbf{p}_v}, m \in \mathcal{I}_v)$, it has the density

$$f_{\underline{k}}^{v|\mathbf{p}_v}(\underline{x}) = \frac{\Gamma(\sum_{m=1}^{|\mathcal{I}_v|} \alpha_{m|\underline{k}}^{v|\mathbf{p}_v})}{\prod_{m=1}^{|\mathcal{I}_v|} \Gamma(\alpha_{m|\underline{k}}^{v|\mathbf{p}_v})} \prod_{m=1}^{|\mathcal{I}_v|} x_m^{\alpha_{m|\underline{k}}^{v|\mathbf{p}_v} - 1} I_{T_{|\mathcal{I}_v|}}(\underline{x}),$$

where $x_{|\mathcal{I}_v|} = 1 - \sum_{m=1}^{|\mathcal{I}_v|-1} x_m$ and $T_{n+1} = \{(x_1, \dots, x_n) \in (0, 1)^n : \sum_{i=1}^n x_i < 1\}$.

Moreover, since we assume that the random vectors $(\mathbf{p}_{m|\underline{k}}^{v|\mathbf{p}_v}, m \in \mathcal{I}_v), \underline{k} \in \mathcal{I}_{\mathbf{p}_v}, v \in V$, are independent the joint density of

$$(\mathbf{p}_{m|\underline{k}}^{v|\mathbf{p}_v}, m \in \mathcal{I}_v, \underline{k} \in \mathcal{I}_{\mathbf{p}_v}, v \in V)$$

has the form

$$f(\underline{x}^{v,\underline{k}}, \underline{k} \in \mathcal{I}_{\mathbf{p}_v}, v \in V) = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}_v}} \frac{\Gamma(\sum_{m \in \mathcal{I}_v} \alpha_{m|\underline{k}}^{v|\mathbf{p}_v})}{\prod_{m \in \mathcal{I}_v} \Gamma(\alpha_{m|\underline{k}}^{v|\mathbf{p}_v})} \prod_{m \in \mathcal{I}_v} (x_m^{v,\underline{k}})^{\alpha_{m|\underline{k}}^{v|\mathbf{p}_v} - 1},$$

where the support is a cartesian product of unit simplexes $\times_{v \in V} T_{|\mathcal{I}_{\mathbf{p}_v}|}^{\times |\mathcal{I}_{\mathbf{p}_v}|}$, that is $\underline{x}^{v,\underline{k}} \in T_{|\mathcal{I}_v|}$ and $x_{|\mathcal{I}_v|}^{v,\underline{k}} = 1 - \sum_{m=1}^{|\mathcal{I}_v|-1} x_m^{v,\underline{k}}, \underline{k} \in \mathcal{I}_{\mathbf{p}_v}, v \in V$.

Now we define a \mathcal{P} -Dirichlet distribution, where \mathcal{P} is a family of DAGs with skeleton G .

Definition 3.2. Let \mathcal{P} be a family of DAGs with skeleton G . The random vector $(\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ associated with the graph G has the \mathcal{P} -Dirichlet distribution iff it has \mathbf{p} -Dirichlet distribution for any $\mathbf{p} \in \mathcal{P}$.

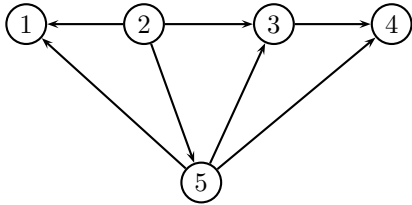
Of course, this definition implies that some consistency conditions for the parameters of the Dirichlet distributions defining the \mathbf{p} -Dirichlet distributions for the various $\mathbf{p} \in \mathcal{P}$, have to be satisfied. This issue is conveniently treated by looking at the moments of the \mathbf{p} -Dirichlet and \mathcal{P} -Dirichlet laws.

Example 3.1. Let $G = (V, E)$ be the decomposable graph with $V = \{1, 2, 3, 4, 5\}$ and cliques $\{1, 2, 5\}$, $\{2, 3, 5\}$ and $\{3, 4, 5\}$. Let $\mathcal{P} = \{\mathbf{p}, \mathbf{p}'\}$ with

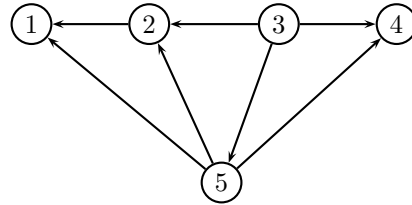
$$\mathbf{p}_1 = \{2, 5\}, \mathbf{p}_2 = \emptyset, \mathbf{p}_3 = \{2, 5\}, \mathbf{p}_4 = \{3, 5\}, \mathbf{p}_5 = \{2\}$$

and

$$\mathbf{p}'_1 = \{2, 5\}, \mathbf{p}'_2 = \{3, 5\}, \mathbf{p}'_3 = \emptyset, \mathbf{p}'_4 = \{3, 5\}, \mathbf{p}'_5 = \{3\}.$$



DAG \mathbf{p} in Example 3.1.



DAG \mathbf{p}' in Example 3.1.

Then the \mathbf{p} -Dirichlet distribution is defined up to a multiplicative constant as a product of the following independent Dirichlets:

$$\begin{aligned} \mathbf{p}^2 &\sim \text{Dir}(\alpha_m^2, m \in \mathcal{I}_2), \quad \mathbf{p}_k^{5|2} \sim \text{Dir}(\alpha_{m|k}^{5|2}, m \in \mathcal{I}_5), \quad k \in \mathcal{I}_2, \quad \mathbf{p}_{\underline{k}}^{1|25} \sim \text{Dir}(\alpha_{m|\underline{k}}^{1|25}, m \in \mathcal{I}_1), \quad \underline{k} \in \mathcal{I}_{25}, \\ \mathbf{p}_{\underline{k}}^{3|25} &\sim \text{Dir}(\alpha_{m|\underline{k}}^{3|25}, m \in \mathcal{I}_3), \quad \underline{k} \in \mathcal{I}_{25}, \quad \mathbf{p}_{\underline{k}}^{4|35} \sim \text{Dir}(\alpha_{m|\underline{k}}^{4|35}, m \in \mathcal{I}_4), \quad \underline{k} \in \mathcal{I}_{35} \end{aligned}$$

and the \mathbf{p}' -Dirichlet is defined through

$$\begin{aligned} \mathbf{p}^3 &\sim \text{Dir}(\beta_m^3, m \in \mathcal{I}_3), \quad \mathbf{p}_k^{5|3} \sim \text{Dir}(\beta_{m|k}^{5|3}, m \in \mathcal{I}_5), \quad k \in \mathcal{I}_3, \quad \mathbf{p}_{\underline{k}}^{1|25} \sim \text{Dir}(\beta_{m|\underline{k}}^{1|25}, m \in \mathcal{I}_1), \quad \underline{k} \in \mathcal{I}_{25}, \\ \mathbf{p}_{\underline{k}}^{2|35} &\sim \text{Dir}(\beta_{m|\underline{k}}^{2|35}, m \in \mathcal{I}_2), \quad \underline{k} \in \mathcal{I}_{35}, \quad \mathbf{p}_{\underline{k}}^{4|35} \sim \text{Dir}(\beta_{m|\underline{k}}^{4|35}, m \in \mathcal{I}_4), \quad \underline{k} \in \mathcal{I}_{35}. \end{aligned}$$

We will come back to this example later in subsection 4.2.1 and see how the constraints on the hyper parameters, i.e., the parameters of the \mathcal{P} -Dirichlet come about.

4. MOMENTS

4.1. The \mathbf{p} -Dirichlet distribution. If the vector of random probabilities $(\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ associated with the graph G follows the \mathbf{p} -Dirichlet distribution, then for any non-negative integers $r_{\underline{i}}, \underline{i} \in \mathcal{I}$,

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{p_v}} \mathbb{E} \prod_{m \in \mathcal{I}_v} [\mathbf{p}_{m|\underline{k}}^{v|p_v}]^{r_{\underline{k},m}^{q_v}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{p_v}} \frac{\Gamma(\sum_{m \in \mathcal{I}_v} \alpha_{m|\underline{k}}^{v|p_v})}{\prod_{m \in \mathcal{I}_v} \Gamma(\alpha_{m|\underline{k}}^{v|p_v})} \frac{\prod_{m \in \mathcal{I}_v} \Gamma(\alpha_{m|\underline{k}}^{v|p_v} + r_{\underline{k},m}^{q_v})}{\Gamma(\sum_{m \in \mathcal{I}_v} \alpha_{m|\underline{k}}^{v|p_v} + r_{\underline{k}}^{p_v})},$$

where

$$r_{\underline{k},m}^{q_v} = \sum_{\underline{i} \in \mathcal{I}: i_{q_v} = (\underline{k}, m)} r_{\underline{i}} \quad \text{and} \quad r_{\underline{k}}^{p_v} = \sum_{\underline{i} \in \mathcal{I}: i_{p_v} = \underline{k}} r_{\underline{i}}.$$

That is

$$(16) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{p_v}} \frac{\prod_{m \in \mathcal{I}_v} (\alpha_{m|\underline{k}}^{v|p_v})^{r_{\underline{k},m}^{q_v}}}{(\tilde{\alpha}_{\underline{k}}^{p_v})^{r_{\underline{k}}^{p_v}}},$$

where we write the rising factorial power $\alpha(\alpha+1)\dots(\alpha+r-1)$ as

$$(\alpha)^r = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)}$$

and

$$(17) \quad \tilde{\alpha}_{\underline{k}}^{p_v} = \sum_{m \in \mathcal{I}_v} \alpha_{m|\underline{k}}^{v|p_v}.$$

Note that since the \mathbf{p} -distribution has a bounded support it is uniquely determined by the moments as given in (16).

4.2. The \mathcal{P} -Dirichlet distribution. Let $G = (V, E)$ be a decomposable graph as above. Denote by \mathcal{C} the set of its cliques and by \mathcal{S} the set of its separators.

If a vector of random probabilities $(\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ associated with the graph G follows the \mathcal{P} -Dirichlet distribution then the formula for moments (16) holds for all $\mathbf{p} \in \mathcal{P}$.

Since its support is bounded, the \mathcal{P} -Dirichlet distribution can be defined through the form of its moments. The equality of the moments for the different representations of the \mathbf{p} -Dirichlet, $p \in \mathcal{P}$, will impose equality constraints on the parameters of the \mathbf{p} -Dirichlet's and consequently those of the \mathcal{P} -Dirichlet. Before developing the theory let us illustrate the mechanism using Example 3.1 above.

4.2.1. Example 3.1 continued. Given nonnegative integers $r_{\underline{i}}, \underline{i} \in \mathcal{I}$, for any $D \subset V$ we define

$$r_{\underline{m}}^D = \sum_{\underline{i} \in \mathcal{I}: \underline{i}_D = \underline{m}} r_{\underline{i}}.$$

For $D = \emptyset$ we write $r^{\emptyset} = r$. The equality of the moments obtained from (16) for both \mathbf{p} and \mathbf{p}' yields

$$\begin{aligned} \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} \mathbf{p}(\underline{i})^{r_{\underline{i}}} &= \frac{\prod_{m \in \mathcal{I}_2} (\alpha_m^2)^{r_m^2}}{(\tilde{\alpha})^r} \prod_{n \in \mathcal{I}_2} \frac{\prod_{m \in \mathcal{I}_5} (\alpha_{(m|n)}^{5|2})^{r_{(n,m)}^{25}}}{(\tilde{\alpha}_n^2)^{r_n^2}} \prod_{\underline{n} \in \mathcal{I}_{25}} \frac{\prod_{m \in \mathcal{I}_1} (\alpha_{(m|\underline{n})}^{1|25})^{r_{(m,\underline{n})}^{125}}}{(\tilde{\alpha}_{\underline{n}}^{25,1})^{r_{\underline{n}}^{25}}} \\ &\quad \times \prod_{\underline{n} \in \mathcal{I}_{25}} \frac{\prod_{m \in \mathcal{I}_3} (\alpha_{(m|\underline{n})}^{3|25})^{r_{(m,\underline{n})}^{325}}}{(\tilde{\alpha}_{\underline{n}}^{25,3})^{r_{\underline{n}}^{25}}} \prod_{\underline{n} \in \mathcal{I}_{35}} \frac{\prod_{m \in \mathcal{I}_4} (\alpha_{(m|\underline{n})}^{4|35})^{r_{(m,\underline{n})}^{435}}}{(\tilde{\alpha}_{\underline{n}}^{35})^{r_{\underline{n}}^{35}}} \\ &= \frac{\prod_{m \in \mathcal{I}_3} (\beta_m^3)^{r_m^3}}{(\tilde{\beta})^r} \prod_{n \in \mathcal{I}_3} \frac{\prod_{m \in \mathcal{I}_5} (\beta_{(m|n)}^{5|3})^{r_{(n,m)}^{35}}}{(\tilde{\beta}_n^3)^{r_n^3}} \prod_{\underline{n} \in \mathcal{I}_{35}} \frac{\prod_{m \in \mathcal{I}_2} (\beta_{(m|\underline{n})}^{2|35})^{r_{(m,\underline{n})}^{235}}}{(\tilde{\beta}_{\underline{n}}^{35,2})^{r_{\underline{n}}^{35}}} \\ &\quad \times \prod_{\underline{n} \in \mathcal{I}_{35}} \frac{\prod_{m \in \mathcal{I}_4} (\beta_{(m|\underline{n})}^{4|35})^{r_{(m,\underline{n})}^{435}}}{(\tilde{\beta}_{\underline{n}}^{35,4})^{r_{\underline{n}}^{35}}} \prod_{\underline{n} \in \mathcal{I}_{25}} \frac{\prod_{m \in \mathcal{I}_1} (\beta_{(m|\underline{n})}^{1|25})^{r_{(m,\underline{n})}^{125}}}{(\tilde{\beta}_{\underline{n}}^{25})^{r_{\underline{n}}^{25}}}. \end{aligned}$$

Since there are no factorial powers in r_m^2 on the right-hand side of the equation above the terms in r_m^2 on the left-hand side must cancel out, that is $\alpha_m^2 = \tilde{\alpha}_m^2$. Similarly, $\beta_m^3 = \tilde{\beta}_m^3$. The factorial power $r_{\underline{n}}^{125}$ on the right- and left-hand side must be the same and therefore $\alpha_{(m,\underline{n})}^{1|25} = \beta_{(m,\underline{n})}^{1|25}$. Similarly, $\alpha_{(m,\underline{n})}^{3|25} = \beta_{(m,\underline{n})}^{2|35}$, $\alpha_{(m,\underline{n})}^{4|35} = \beta_{(m,\underline{n})}^{4|35}$ and also $\tilde{\alpha} = \tilde{\beta}$. For the factorial powers in $r_{\underline{n}}^{25}$ we observe that on the left-hand side there is one power in the numerator and two in the denominator, while on the right-hand side there is only one power in the denominator. Therefore the factorial power in the numerator must cancel with one of the two factorial powers of $\tilde{\alpha}_{\underline{n}}^{25,3}$ or of $\tilde{\alpha}_{\underline{n}}^{25,1}$ in the denominator. This means that

- either we have the cancelation $\alpha_{\underline{n}}^{5|2} = \tilde{\alpha}_{\underline{n}}^{25,3}$ and therefore $\tilde{\beta}_{\underline{n}}^{25} = \tilde{\alpha}_{\underline{n}}^{25,1}, \forall \underline{n} \in \mathcal{I}_{25}$
- or we have the cancelation $\alpha_{\underline{n}}^{5|2} = \tilde{\alpha}_{\underline{n}}^{25,1}$ and therefore $\tilde{\beta}_{\underline{n}}^{25} = \tilde{\alpha}_{\underline{n}}^{25,3}, \forall \underline{n} \in \mathcal{I}_{25}$.

The first choice means that we associate the separator $\{2, 5\}$ with the clique $\{1, 2, 5\}$ while in the second we associate $\{2, 5\}$ with the clique $\{2, 3, 5\}$. This two choices correspond to two different \mathbf{p} -perfect orders of the cliques:

$$(18) \quad o_{\mathbf{p}}^{(1)} = (C_1 = \{2, 3, 5\}, C_2 = \{1, 2, 5\}, C_3 = \{3, 4, 5\})$$

$$(19) \quad o_{\mathbf{p}}^{(2)} = (C_1 = \{1, 2, 5\}, C_2 = \{2, 3, 5\}, C_3 = \{3, 4, 5\}),$$

respectively. Of course, we could also exchange the cliques C_2 and C_3 in both orders. What is important is the pairings $(\{2, 5\}, \{1, 2, 5\})$ or $(\{2, 5\}, \{2, 3, 5\})$, respectively.

Similarly, for the factorial powers $r_{\underline{n}}^{35}$ on the right-hand side one can choose to cancel the factorial power of $r_{\underline{n}}^{35}$ in the numerator with factorial powers of either $\tilde{\beta}_{\underline{n}}^{35,2}$ or $\tilde{\beta}_{\underline{n}}^{35,4}$. Consequently,

$$\text{either } \beta_{\underline{n}}^{5|3} = \tilde{\beta}_{\underline{n}}^{35,2} \text{ and therefore } \tilde{\alpha}_{\underline{n}}^{35} = \tilde{\beta}_{\underline{n}}^{35,4} \quad \forall \underline{n} \in \mathcal{I}_{35},$$

$$\text{or } \beta_{\underline{n}}^{5|3} = \tilde{\beta}_{\underline{n}}^{35,4} \text{ and therefore } \tilde{\alpha}_{\underline{n}}^{35} = \tilde{\alpha}_{\underline{n}}^{35,2} \quad \forall \underline{n} \in \mathcal{I}_{35},$$

which corresponds to the two \mathbf{p}' -perfect orders:

$$(20) \quad o_{\mathbf{p}'}^{(1)} = (C_1 = \{2, 3, 5\}, C_2 = \{3, 4, 5\}, C_3 = \{1, 2, 5\})$$

$$(21) \quad o_{\mathbf{p}'}^{(2)} = (C_1 = \{3, 4, 5\}, C_2 = \{2, 3, 5\}, C_3 = \{1, 2, 5\}),$$

respectively. Again here we could exchange C_2 and C_3 in both cases. What is important are the pairings $(\{3, 5\}, \{3, 4, 5\})$ and $(\{3, 5\}, \{2, 3, 5\})$, respectively.

From any of these cancelation possibilities we obtain the same formula of moments

$$\mathbb{E} \prod_{i \in \mathcal{I}} \mathbf{p}(\underline{i})^{r_i} = \frac{\prod_{\underline{n} \in \mathcal{I}_{125}} (\nu_{\underline{n}}^{125})^{r_{\underline{n}}^{125}} \prod_{\underline{n} \in \mathcal{I}_{235}} (\nu_{\underline{n}}^{235})^{r_{\underline{n}}^{235}} \prod_{\underline{n} \in \mathcal{I}_{345}} (\nu_{\underline{n}}^{345})^{r_{\underline{n}}^{345}}}{(\mu)^r \prod_{\underline{m} \in \mathcal{I}_{25}} (\mu_{\underline{m}}^{25})^{r_{\underline{m}}^{25}} \prod_{\underline{m} \in \mathcal{I}_{35}} (\mu_{\underline{m}}^{35})^{r_{\underline{m}}^{35}}},$$

but with different constraints for the parameters:

either (I)

$$\begin{aligned} \mu &= \sum_{\underline{m} \in \mathcal{I}_{125}} \nu_{\underline{m}}^{125} = \sum_{\underline{m} \in \mathcal{I}_{345}} \nu_{\underline{m}}^{345} (= \sum_{\underline{m} \in \mathcal{I}_{235}} \nu_{\underline{m}}^{235}), \\ \mu_{\underline{n}}^{25} &= \sum_{m \in \mathcal{I}_1} \nu_{(m, \underline{n})}^{125} = \sum_{m \in \mathcal{I}_3} \nu_{(m, \underline{n})}^{235}, \\ \mu_{\underline{n}}^{35} &= \sum_{m \in \mathcal{I}_4} \nu_{(m, \underline{n})}^{345} = \sum_{m \in \mathcal{I}_2} \nu_{(m, \underline{n})}^{235}, \end{aligned}$$

corresponding to the family of orders $O_{\mathcal{P}} = (o_{\mathbf{p}}^{(2)}, o_{\mathbf{p}'}^{(2)})$.

or (II)

$$\begin{aligned} \mu &= \sum_{\underline{m} \in \mathcal{I}_{125}} \nu_{\underline{m}}^{125} = \sum_{\underline{m} \in \mathcal{I}_{235}} \nu_{\underline{m}}^{235}, \\ \mu_{\underline{n}}^{25} &= \sum_{m \in \mathcal{I}_1} \nu_{(m, \underline{n})}^{125} = \sum_{m \in \mathcal{I}_3} \nu_{(m, \underline{n})}^{235}, \\ \mu_{\underline{n}}^{35} &= \sum_{m \in \mathcal{I}_4} \nu_{(m, \underline{n})}^{345}, \end{aligned}$$

corresponding to the family of orders $O_{\mathcal{P}} = (o_{\mathbf{p}}^{(2)}, o_{\mathbf{p}'}^{(1)})$.
or (III)

$$\begin{aligned}\mu &= \sum_{\underline{m} \in \mathcal{I}_{235}} \nu_{\underline{m}}^{235} = \sum_{\underline{m} \in \mathcal{I}_{345}} \nu_{\underline{m}}^{345}, \\ \mu_{\underline{n}}^{25} &= \sum_{m \in \mathcal{I}_1} \nu_{(m, \underline{n})}^{125}, \\ \mu_{\underline{n}}^{35} &= \sum_{m \in \mathcal{I}_4} \nu_{(m, \underline{n})}^{345} = \sum_{m \in \mathcal{I}_2} \nu_{(m, \underline{n})}^{235},\end{aligned}$$

corresponding to the family of orders $O_{\mathcal{P}} = (o_{\mathbf{p}}^{(1)}, o_{\mathbf{p}'}^{(2)})$.
or (IV)

$$\begin{aligned}\mu &= \sum_{\underline{m} \in \mathcal{I}_{235}} \nu_{\underline{m}}^{235}, \\ \mu_{\underline{n}}^{25} &= \sum_{m \in \mathcal{I}_1} \nu_{(m, \underline{n})}^{125}, \\ \mu_{\underline{n}}^{35} &= \sum_{m \in \mathcal{I}_4} \nu_{(m, \underline{n})}^{345},\end{aligned}$$

corresponding to the family of orders $O_{\mathcal{P}} = (o_{\mathbf{p}}^{(1)}, o_{\mathbf{p}'}^{(1)})$.

We note that we obtained four different families of distributions, that is as many as the number of combinations of pairs (S_l, C_l) , where $S_l = \{2, 5\}$ and $\{3, 5\}$ here. Of course, choices will multiply with the number of separators with different possible pairings. In fact, more generally, choices may multiply with the number of elements of \mathfrak{P} with different possible pairings in \mathfrak{Q} (see definitions (23) below) and also with the size of \mathcal{P} .

In this example, we see that we have the poset $(IV) \rightarrow (II, III) \rightarrow (I)$ of families of \mathcal{P} -Dirichlet distributions, with family (IV) being the maximal family.

In the remainder of this section we will show that to each collection of orders

$$(22) \quad O_{\mathcal{P}} = (o_{\mathbf{p}}, \mathbf{p} \in \mathcal{P} : o_{\mathbf{p}} \text{ is } \mathbf{p} - \text{perfect}, \mathbf{p} \in \mathcal{P})$$

corresponds a family of \mathcal{P} -Dirichlet distributions. Though in our example it is easy to see that (IV) is the maximal family of the poset we are unable to prove that for any given \mathcal{P} , there exists such a unique maximal family.

4.2.2. The moment formula. To find a convenient expression for the moment formula we need the following two auxiliary results.

Lemma 4.1. *Consider a DAG with skeleton G defined by a parent function \mathbf{p} . Then,*

$$\mathcal{C} \subset \mathbf{q}(V) \quad \text{and} \quad \mathcal{S} \subset \mathbf{p}(V).$$

Moreover,

$$\mathbf{p}(V) \setminus \mathcal{S} = \mathbf{q}(V) \setminus \mathcal{C}.$$

Proof. Let $o = (C_1, \dots, C_K)$ be a \mathbf{p} -perfect order of cliques. Using the notation $v_{l,j}$, $j = s_l + 1, \dots, c_l$, $l = 1, \dots, K$, introduced in (3), we see that $\mathbf{p}_{v_{l,s_l+1}} = S_l$ and $\mathbf{q}_{v_{l,c_l}} = C_l$, $l = 1, \dots, K$ (here we set $S_1 = \emptyset$) so that the first statement of the lemma is proved.

For the second part note that for any $l = 1, \dots, K$ and any $j = s_l + 2, \dots, c_l$, we have $\mathbf{q}_{v_{l,j-1}} = \mathbf{p}_{v_{l,j}} \notin \mathcal{C} \cup \mathcal{S}$. That is

$$\mathbf{p}(V) \setminus \mathcal{S} = \mathbf{q}(V) \setminus \mathcal{C} = \bigcup_{l=1}^K \bigcup_{j=s_l+2}^{c_l} \mathbf{p}_{v_{l,j}}.$$

□

Note that if $v, w \in V$ are distinct then \mathbf{q}_v and \mathbf{q}_w are also distinct, but \mathbf{p}_v and \mathbf{p}_w may be the same. It means that while considering the set $\mathbf{p}(V)$ we allow for a given set to appear in $\mathbf{p}(V)$ more than once, that is, each element of $\mathbf{p}(V)$ has its multiplicity, e.g. a set S which is a separator can happen to be a multiple separator, moreover such a set S can appear, at most once (due to Lemma 4.1), in $\mathbf{p}(V) \setminus \mathcal{S}$.

Denote

$$\mathfrak{R}_{\mathbf{p}} := \mathbf{p}(V) \setminus \mathcal{S} = \mathbf{q}(V) \setminus \mathcal{C}.$$

Let \mathcal{P} be a family of DAGs with the same skeleton G . Since in a DAG in \mathcal{P} there are no immoralities, they are all Markov equivalent. Consider the following sets

$$(23) \quad \Omega = \bigcap_{\mathbf{p} \in \mathcal{P}} \mathbf{q}(V) \supset \mathcal{C}, \quad \text{and} \quad \mathfrak{P} = \bigcap_{\mathbf{p} \in \mathcal{P}} \mathbf{p}(V) \supset \mathcal{S},$$

where in the definition of \mathfrak{P} we allow for multiple separators. By Lemma 4.1 it follows that

$$\Omega = \mathcal{C} \cup \mathfrak{R} \quad \text{and} \quad \mathfrak{P} = \mathcal{S} \cup \mathfrak{R},$$

where $\mathfrak{R} = \bigcap_{\mathbf{p} \in \mathcal{P}} \mathfrak{R}_{\mathbf{p}}$.

Lemma 4.2. *For a family \mathcal{P} of DAGs with skeleton G , let $O_{\mathcal{P}}$ be a collection of \mathbf{p} -perfect orders of cliques, $\mathbf{p} \in \mathcal{P}$. For any clique C , let \mathfrak{R}_C denote a family of these elements of \mathfrak{R} which are contained in C . If there exists $o \in O_{\mathcal{P}}$ such that $\mathcal{S} \ni S \xrightarrow{o} C$ then all the elements $Q_1^C, \dots, Q_{j_C-1}^C \in \mathfrak{R}_C$ (it may be empty) can be numbered as follows*

$$(24) \quad S =: Q_{j_C}^C \subsetneq Q_{j_C-1}^C \subsetneq Q_{j_C-2}^C \subsetneq \dots \subsetneq Q_2^C \subsetneq Q_1^C \subsetneq Q_0^C := C.$$

Proof. Let o be \mathbf{p} -perfect for a $\mathbf{p} \in \mathcal{P}$. The result follows immediately from the fact (see (3) and the proof of Lemma 4.1) that possible sets Q_i^C , $1 \leq i \leq j_C - 1$, from \mathfrak{R}_C are of the form $\mathbf{q}_{v_{l,k}} = \mathbf{p}_{v_{l,k+1}}$, $k = s_l + 1, \dots, c_l - 1$, moreover $S = \mathbf{p}_{v_{l,s_l+1}} \subsetneq \mathbf{q}_{v_{l,s_l+1}}$ and $C = \mathbf{q}_{v_{l,c_l}} \supsetneq \mathbf{q}_{v_{l,c_l-1}}$, $l = 1, \dots, K$, □

Now we are in a position to give the formula for moments.

Theorem 4.3. *A vector of random probabilities $(\mathbf{p}(\underline{i}), i \in \mathcal{I})$ associated with the graph G has a \mathcal{P} -Dirichlet distribution iff there exists a collection $O_{\mathcal{P}}$ of \mathbf{p} -perfect orders, $\mathbf{p} \in \mathcal{P}$, as in (22), such that for any $r_{\underline{i}} \in \mathbb{N}$, $\underline{i} \in \mathcal{I}$,*

$$(25) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} (\mathbf{p}(\underline{i}))^{r_{\underline{i}}} = \frac{\prod_{A \in \Omega} \prod_{\underline{m} \in \mathcal{I}_A} (\nu_{\underline{m}}^A)^{r_{\underline{m}}^A}}{\prod_{B \in \mathfrak{P}} \prod_{\underline{n} \in \mathcal{I}_B} (\mu_{\underline{n}}^B)^{r_{\underline{n}}^B}}$$

where $\nu_{\underline{m}}^A$, $\underline{m} \in \mathcal{I}_A$, $A \in \Omega$, and $\mu_{\underline{n}}^B$, $\underline{n} \in \mathcal{I}_B$, $B \in \mathfrak{P}$, are positive numbers satisfying

$$(26) \quad \mu_{\underline{n}}^B = \sum_{\underline{k} \in \mathcal{I}_{A \setminus B}} \nu_{(\underline{n}, \underline{k})}^A \quad \forall \underline{n} \in \mathcal{I}_B$$

whenever there exist $\mathcal{S} \ni S \subset C \in \mathcal{C}$ and $o \in O_{\mathcal{P}}$ such that $S \xrightarrow{o} C$ and $B = Q_i^C \subsetneq A = Q_{i-1}^C$ for some $i \in \{1, \dots, j_C\}$.

The proof is given in the Appendix.

It follows from the expression (25) of the moments that the \mathcal{P} -Dirichlet distribution has a density which, following a given perfect order $o \in O_{\mathcal{P}}$ of the cliques, can be expressed as the product of independent classical Dirichlet distributions as given below. Using the notation $R_i^{C_l} = Q_i^{C_l} \setminus Q_{i+1}^{C_l}$, $i \in \{0, \dots, j_{C_l} - 1\}$, $l \in \{1, \dots, K\}$, we have the following.

Corollary 4.4. *Let $(\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ be a random vector having \mathcal{P} -Dirichlet distribution with constraints (26) governed by a family $O_{\mathcal{P}}$. Consider any perfect order $o = (C_1, \dots, C_K) \in O_{\mathcal{P}}$ of the cliques. There exist independent vectors $(\mathbf{p}_{\underline{m}|\underline{k}}^{R_i^{C_l}|Q_{i+1}^{C_l}}, \underline{m} \in \mathcal{I}_{R_i^{C_l}})$ having classical Dirichlet distributions $\text{Dir}(\nu_{(\underline{k}, \underline{m})}^{Q_i^{C_l}}, \underline{m} \in \mathcal{I}_{R_i^{C_l}})$, $\underline{k} \in \mathcal{I}_{Q_{i+1}^{C_l}}$, $i \in \{0, 1, \dots, j_{C_l} - 1\}$, $l \in \{1, \dots, K\}$, such that for any $\underline{i} \in \mathcal{I}$*

$$\mathbf{p}(\underline{i}) = \prod_{l=1}^K \prod_{i=0}^{j_{C_l}-1} \mathbf{p}_{\underline{m}|\underline{k}}^{R_i^{C_l}|Q_{i+1}^{C_l}}, \quad \text{where } \underline{m} = \underline{i}_{R_i^{C_l}} \text{ and } \underline{k} = \underline{i}_{Q_{i+1}^{C_l}}.$$

Thus the density of the \mathcal{P} -Dirichlet distribution can be written as

$$(27) \quad \prod_{l=1}^K \prod_{i=0}^{j_{C_l}-1} \prod_{\underline{k} \in \mathcal{I}_{Q_{i+1}^{C_l}}} \text{Dir}(\nu_{(\underline{k}, \underline{m})}^{Q_i^{C_l}}, \underline{m} \in \mathcal{I}_{R_i^{C_l}}).$$

We note that the well-known decomposition of the hyper Dirichlet density (10) into

$$C \prod_{l=1}^K \prod_{\underline{k} \in \mathcal{I}_{S_l}} \prod_{\underline{m} \in \mathcal{I}_{R_l}} (\mathbf{p}_{\underline{m}|\underline{k}}^{R_l|S_l})^{\alpha_{(\underline{m}, \underline{k})}^{C_l}}^{-1},$$

where C is the normalizing constant, is a special case of (27) where $j_c = 1$ and $Q_{i+1}^{C_l}$ and $Q_i^{C_l}$ are respectively replaced by S_l and $R_l = C_l \setminus S_l$.

4.2.3. More examples and the hyper-Dirichlet as a special case of the \mathcal{P} -Dirichlet.

We now give a few examples of the \mathcal{P} -Dirichlet distribution. We start with an example in which $\mathfrak{R} \neq \emptyset$.

Example 4.1. Let $G = (V, E)$ be a graph with

$$V = \{1, 2, 3, 4, 5\} \quad \text{and} \quad E = \{\{1, 3\}, \{2, 4\}, \{3, 4\}, \{3, 5\}, \{4, 5\}\}.$$

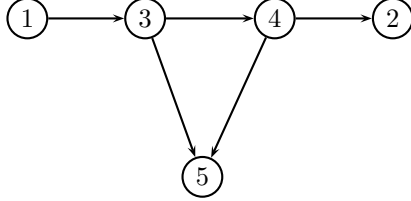
Then $\mathcal{C} = \{\{1, 3\}, \{3, 4, 5\}, \{2, 4\}\}$ and $\mathcal{S} = \{\emptyset, \{3\}, \{4\}\}$.

Let $\mathfrak{P} = \{\mathbf{p}, \mathbf{p}'\}$, where

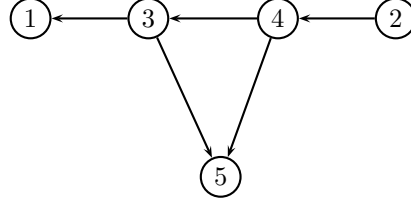
$$\mathbf{p}_1 = \emptyset, \quad \mathbf{p}_2 = \{4\}, \quad \mathbf{p}_3 = \{1\}, \quad \mathbf{p}_4 = \{3\}, \quad \mathbf{p}_5 = \{3, 4\}$$

and

$$\mathbf{p}'_1 = \{3\}, \quad \mathbf{p}'_2 = \emptyset, \quad \mathbf{p}'_3 = \{4\}, \quad \mathbf{p}'_4 = \{2\}, \quad \mathbf{p}'_5 = \{3, 4\}.$$



DAG \mathbf{p} for Example 4.1.



DAG \mathbf{p}' for Example 4.1.

Then

$$\mathfrak{Q} = \mathcal{C} \cup \{3, 4\}, \quad \mathfrak{P} = \mathcal{S} \cup \{3, 4\},$$

$$\mathfrak{R}_{\{3,4,5\}} = \{3, 4\}, \quad \mathfrak{R}_{\{1,3\}} = \mathfrak{R}_{\{2,4\}} = \emptyset.$$

Moreover, there is only one available collection of orders $O_{\mathcal{P}} = \{o, o'\}$, where the \mathbf{p} -perfect order o is the following $C_1 = \{1, 3\}$, $C_2 = \{3, 4, 5\}$, $C_3 = \{2, 4\}$ and the \mathbf{p}' -perfect order o' is the following $C'_1 = \{2, 4\}$, $C'_2 = \{3, 4, 5\}$, $C'_3 = \{1, 3\}$.

Then formula (25) for moments becomes

$$\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r_{\underline{i}}} = \frac{\prod_{\underline{m} \in \mathcal{I}_{13}} (\nu_{\underline{m}}^{13})^{r_{\underline{m}}} \prod_{\underline{m} \in \mathcal{I}_{24}} (\nu_{\underline{m}}^{24})^{r_{\underline{m}}} \prod_{\underline{m} \in \mathcal{I}_{345}} (\nu_{\underline{m}}^{345})^{r_{\underline{m}}} \prod_{\underline{m} \in \mathcal{I}_{34}} (\nu_{\underline{m}}^{34})^{r_{\underline{m}}}}{(\mu)^r \prod_{\underline{m} \in \mathcal{I}_3} (\mu_{\underline{m}}^3)^{r_{\underline{m}}} \prod_{\underline{m} \in \mathcal{I}_4} (\mu_{\underline{m}}^4)^{r_{\underline{m}}} \prod_{\underline{m} \in \mathcal{I}_{34}} (\mu_{\underline{m}}^{34})^{r_{\underline{m}}}}$$

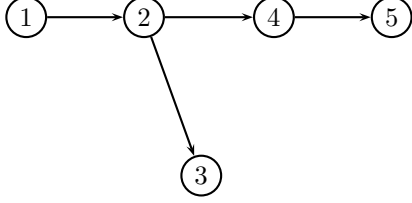
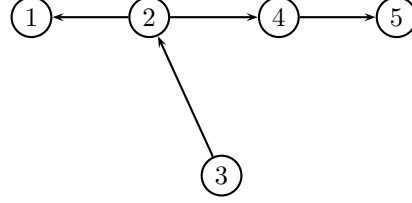
with the following consistency conditions:

$$\begin{aligned} \mu &= \sum_{\underline{k} \in \mathcal{I}_{13}} \nu_{\underline{k}}^{13} = \sum_{\underline{k} \in \mathcal{I}_{24}} \nu_{\underline{k}}^{24}, \\ \mu_n^3 &= \sum_{k \in \mathcal{I}_4} \nu_{(n,k)}^{34} = \sum_{k \in \mathcal{I}_1} \nu_{(k,n)}^{13}, \quad n \in \mathcal{I}_3; \\ \mu_n^4 &= \sum_{k \in \mathcal{I}_3} \nu_{(k,n)}^{34} = \sum_{k \in \mathcal{I}_2} \nu_{(k,n)}^{24}, \quad n \in \mathcal{I}_4; \\ \mu_{\underline{n}}^{34} &= \sum_{k \in \mathcal{I}_5} \nu_{(\underline{n},k)}^{345}, \quad \underline{n} \in \mathcal{I}_{34}. \end{aligned}$$

Consequently, combining the above equations, we also get

$$\mu = \sum_{n \in \mathcal{I}_3} \mu_n^3 = \sum_{n \in \mathcal{I}_4} \mu_n^4 = \sum_{\underline{k} \in \mathcal{I}_{34}} \nu_{\underline{k}}^{34}.$$

We will now look at the case where $\mathfrak{R} = \emptyset$. We will see that the formula for moments simplifies and becomes closer to the moment formula for the hyper Dirichlet. Let us first consider an example.

DAG \mathbf{p} for Example 4.2.DAG \mathbf{p}' for Example 4.2.

Example 4.2. Consider a tree $G = (V, E)$ with $V = \{1, 2, 3, 4, 5\}$ and $E = \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{4, 5\}\}$. Let $\mathcal{P} = \{\mathbf{p}, \mathbf{p}'\}$, where

$$\mathbf{p}_1 = \emptyset, \mathbf{p}_2 = \{1\}, \mathbf{p}_3 = \{2\}, \mathbf{p}_4 = \{2\}, \mathbf{p}_5 = \{4\}$$

and

$$\mathbf{p}'_1 = \{2\}, \mathbf{p}'_2 = \{3\}, \mathbf{p}'_3 = \emptyset, \mathbf{p}'_4 = \{2\}, \mathbf{p}'_5 = \{4\}.$$

Then

$$\mathfrak{Q} = \mathcal{C} = E, \quad \mathfrak{P} = \{\{2\}, \{2\}, \{4\}\} = \mathcal{S}.$$

Moreover $\mathfrak{R}_C = \emptyset$ for any $C \in \mathcal{C}$. For $S = \{4\}$ and $C = \{2, 4\}$ it never happens that $S = S_l$ and $C = C_l$ for perfect orders of cliques C_1, C_2, C_3, C_4 generated by \mathbf{p} or \mathbf{p}' .

That is the formula for moments as in (25) reads

$$\mathbb{E} \prod_{i \in \mathcal{I}} [\mathbf{p}(i)]^{r_i} = \frac{\prod_{m \in \mathcal{I}_{12}} (\nu_m^{12})^{r_m} \prod_{m \in \mathcal{I}_{23}} (\nu_m^{23})^{r_m} \prod_{m \in \mathcal{I}_{24}} (\nu_m^{24})^{r_m} \prod_{m \in \mathcal{I}_{45}} (\nu_m^{45})^{r_m}}{(\mu)^r \prod_{m \in \mathcal{I}_2} (\mu_m^2)^{2r_m} \prod_{m \in \mathcal{I}_4} (\mu_m^4)^{r_m}},$$

with the following consistency conditions: either we consider the set of orders: \mathbf{p} -perfect o of the form $C_1 = \{1, 2\}$, $C_2 = \{2, 3\}$, $C_3 = \{2, 4\}$, $C_4 = \{4, 5\}$ and \mathbf{p}' -perfect o' of the form $C'_1 = \{2, 3\}$, $C'_2 = \{1, 2\}$, $C'_3 = \{2, 4\}$, $C'_4 = \{4, 5\}$ and then the constraints

$$\begin{aligned} \mu &= \sum_{\underline{k} \in \mathcal{I}_{12}} \nu_{\underline{k}}^{12} = \sum_{\underline{k} \in \mathcal{I}_{23}} \nu_{\underline{k}}^{23}, \\ \mu_n^{2(1)} &= \sum_{k \in \mathcal{I}_1} \nu_{(k,n)}^{12} = \sum_{k \in \mathcal{I}_3} \nu_{(n,k)}^{23}, \quad n \in \mathcal{I}_2, \\ \mu_n^{2(2)} &= \sum_{k \in \mathcal{I}_4} \nu_{(n,k)}^{24}, \quad n \in \mathcal{I}_2, \\ \mu_n^4 &= \sum_{k \in \mathcal{I}_5} \nu_{(n,k)}^{45}, \quad n \in \mathcal{I}_4. \end{aligned}$$

or we keep o as defined above and for o' we take $C'_1 = \{2, 3\}$, $C'_2 = \{2, 4\}$, $C'_3 = \{4, 5\}$, $C'_4 = \{1, 2\}$ and the constraints are more restrictive since while the first and the fourth line above are preserved the second and the third merge into

$$\mu_n^{2(1)} = \mu_n^{2(2)} = \sum_{k \in \mathcal{I}_1} \nu_{(k,n)}^{12} = \sum_{k \in \mathcal{I}_3} \nu_{(n,k)}^{23} = \sum_{k \in \mathcal{I}_4} \nu_{(n,k)}^{24}, \quad n \in \mathcal{I}_2.$$

The above example falls under a more general setting which follows immediately from Theorem 4.3 and which we formalize as:

Corollary 4.5. *Let a random vector $(\mathbf{p}(\underline{i}))_{\underline{i} \in \mathcal{I}}$ associated with the graph G have a \mathcal{P} -Dirichlet distribution for a family \mathcal{P} of moral DAGs and a collection $O_{\mathcal{P}}$ of \mathbf{p} -perfect orders, $\mathbf{p} \in \mathcal{P}$. If \mathcal{P} has the property $\mathfrak{Q} = \mathcal{C}$ (that is, $\mathfrak{P} = \mathcal{S}$ or $\mathfrak{R} = \emptyset$) then*

$$(28) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r_{\underline{i}}} = \frac{\prod_{C \in \mathcal{C}} \prod_{\underline{m} \in \mathcal{I}_C} (\nu_{\underline{m}}^C)^{r_{\underline{m}}^C}}{(\mu)^r \prod_{S \in \mathcal{S}} \prod_{\underline{m} \in \mathcal{I}_S} (\mu_{\underline{m}}^S)^{r_{\underline{m}}^S}},$$

with the following consistency conditions

$$(29) \quad \sum_{\underline{m} \in \mathcal{C}: \underline{m}_S = \underline{n}} \nu_{\underline{m}}^C = \mu_{\underline{n}}^S, \quad \underline{n} \in \mathcal{I}_S$$

whenever $S \xrightarrow{o} C$ for some $o \in O_{\mathcal{P}}$.

The hyper-Dirichlet distribution is, of course, uniquely defined through moments of the form exactly the same as in (28) but the consistency conditions are even stronger: (29) is satisfied for any $\mathcal{S} \ni S \subset C \in \mathcal{C}$, where $S = \emptyset \in \mathcal{S}$. Therefore a \mathcal{P} -Dirichlet distribution for the family \mathcal{P} of *all* moral DAGs with skeleton $G = (V, E)$ is a hyper-Dirichlet distribution. Actually, it follows directly from Corollary 4.5 above that a considerably smaller family \mathcal{P} forces the \mathcal{P} -Dirichlet to be hyper-Dirichlet. We state this as follows.

Theorem 4.6. *Let \mathcal{L} be a \mathcal{P} -Dirichlet distribution for a family \mathcal{P} of moral DAGs and a collection $O_{\mathcal{P}}$ of \mathbf{p} -perfect orders, $\mathbf{p} \in \mathcal{P}$. If*

$$(30) \quad \mathfrak{R} = \emptyset$$

and

$$(31) \quad \forall (S \in \mathcal{S}, C \in \mathcal{C}) \text{ if } S \subset C \text{ then } \exists o \in O_{\mathcal{P}} \text{ such that } S \xrightarrow{o} C$$

then \mathcal{L} is a hyper-Dirichlet distribution.

5. THE \mathcal{P} -DIRICHLET AS A PRIOR DISTRIBUTION

In this section, we look at the properties of the \mathcal{P} -Dirichlet as a prior distribution. We first compute the dimension of the \mathcal{P} -Dirichlet family for a given \mathcal{P} and show that it is always larger than the dimension of the hyper Dirichlet family with the same skeleton G (unless the set \mathcal{P} is so large that the corresponding \mathfrak{R} is empty and (31) holds, of course). We then show that the \mathcal{P} -Dirichlet is a conjugate distribution with the directed strong hyper Markov property for every $\mathbf{p} \in \mathcal{P}$. This, of course, makes the \mathcal{P} -Dirichlet easy to use in a Bayesian model selection process. In Section 5.3, we argue that, when we have additional constraints on the direction of certain edges between variables, the \mathcal{P} -Dirichlet arises naturally as a conjugate prior when doing model selection in the class of models Markov with respect to a DAG with a given decomposable skeleton G .

5.1. Dimension of the \mathcal{P} -Dirichlet family. We are now going to show that the dimension of the parameter space of the \mathcal{P} -Dirichlet distribution is always greater (or equal) than that of the hyper Dirichlet. This means, of course, that when choosing the \mathcal{P} -Dirichlet as a prior rather than the hyper Dirichlet, we gain flexibility in our choice of the hyper parameters. The dimensions of both families are given in the following theorem.

Let G be a decomposable graph with \mathcal{C} as its set of cliques and \mathcal{S} as its set of separators. If \mathcal{P} is a collection of DAG's with skeleton G and $O_{\mathcal{P}}$ a collection of \mathfrak{p} -perfect orders, $\mathfrak{p} \in \mathcal{P}$, for $S \in \mathcal{S}$ given, we denote by N_S be the number of cliques C such that

$$\text{if } S \subset C \text{ then } \exists o \in O_{\mathcal{P}} \text{ such that } S \xrightarrow{o} C.$$

Recall for that for $v \in V$, \mathcal{I}_v denotes the set of values that X_v can take and $|\mathcal{I}_v|$ denotes the cardinality of \mathcal{I}_v .

Theorem 5.1. *For G and \mathcal{P} as given above, the dimension of the parameter space of the \mathcal{P} -Dirichlet family of distributions is*

$$(32) \quad \mathcal{N}_{\mathcal{P}} = \sum_{Q \in \Omega} \prod_{v \in Q} |\mathcal{I}_v| - \sum_{S \in \mathcal{S}} (N_S - 1) \prod_{v \in S} |\mathcal{I}_v|.$$

The dimension of the parameter space of the hyper Dirichlet family of distributions with the same skeleton G is equal to

$$(33) \quad \mathcal{N}_{HP} = \sum_{C \in \mathcal{C}} \prod_{v \in C} |\mathcal{I}_v| - \sum_{S \in \mathcal{S}} (N_S - 1) \prod_{v \in S} |\mathcal{I}_v|.$$

Moreover, if the \mathcal{P} -Dirichlet is not identical to the hyper Dirichlet, we always have

$$(34) \quad \mathcal{N}_{\mathcal{P}} > \mathcal{N}_{HP}.$$

Proof. From (25) and (26), we see that the parameters are the $\nu_{\underline{m}}^A$ and we need not count the $\mu_{\underline{n}}^B$ since they are defined by the constraints of the type (26). There clearly are $\sum_{Q \in \Omega} \prod_{v \in Q} |\mathcal{I}_v|$ such parameters $\nu_{\underline{m}}^A$. They are not all free since an element $S \in \mathcal{S}$ can be equal to an element $Q_{j_C}^C$ for several $C \in \mathcal{C}$. More precisely for all $C \in \mathcal{C}$ such that there exists $o \in O_{\mathcal{P}}$ with $S \xrightarrow{o} C$, we would have

$$Q_{j_C}^C \subsetneq Q_{j_{C-1}}^C$$

and therefore by (26), we have $(N_S - 1)$ equality of the type

$$\mu_{\underline{n}}^S = \mu_{\underline{n}}^{Q_{j_C}^C} = \sum_{\underline{k} \in I_{Q_{j_{C-1}}^C} \setminus S} \nu_{(\underline{n}, \underline{k})}^{Q_{j_{C-1}}^C}$$

and thus $(N_S - 1)$ constraints for a given $\nu_{(\underline{n}, \underline{k})}^{Q_{j_{C-1}}^C}$ and thus a total of $(N_S - 1) \prod_{v \in S} |\mathcal{I}_v|$ constraints for each S . We now note that if $B \in \mathfrak{R}$, that is if B is not a separator, the corresponding equation (26) is not a constraint since then there is only one clique to which B can belong, i.e. only one A such that

$$B = Q_i^C \subsetneq A = Q_{i-1}^C \text{ and } \mu_{\underline{n}}^B = \sum_{\underline{k} \in I_{A \setminus B}} \nu_{(\underline{n}, \underline{k})}^A.$$

It follows that (32) is proved.

In the case of the hyper Dirichlet, a similar argument shows us that the total number of parameters is $\sum_{C \in \mathcal{C}} \prod_{v \in Q} |\mathcal{I}_v|$. The constraints given by equations of the type (26) are of the form

$$\mu_{\underline{n}}^S = \sum_{\underline{k} \in C \setminus S} \nu_{(\underline{n}, \underline{k})}^C$$

for any C containing S . Since considering the hyper Dirichlet is equivalent to taking \mathcal{P} as the set of all DAG's Markov equivalent to G , N_S is nothing but the number of cliques containing S and equation (33) follows.

To see that $\mathcal{N}_{\mathcal{P}}$ is always strictly greater than \mathcal{N}_{HP} , we observe that unless $\mathfrak{R} = \emptyset$, that is the \mathcal{P} -Dirichlet is the hyper Dirichlet, $\mathcal{C} \subsetneq \mathfrak{Q}$ and therefore

$$\sum_{Q \in \mathfrak{Q}} \prod_{v \in Q} |\mathcal{I}_v| > \sum_{C \in \mathcal{C}} \prod_{v \in C} |\mathcal{I}_v|$$

Moreover, for each S , N_S in the \mathcal{P} -Dirichlet is less than or equal to the corresponding N_S in the hyper Dirichlet. Inequality (34) follows immediately. \square

Let us illustrate this result by deriving the dimension of \mathcal{P} -Dirichlet and the hyper Dirichlet families respectively when G and \mathcal{P} are as given in Example 4.1.

Example 5.1. (Example 4.1 continued) Let us assume that all variables are binary, that is $|\mathcal{I}_v| = 2$, $v = 1, 2, 3, 4, 5$. From the expression of the moments in (25), we see that the number of $\nu_{\underline{m}}^A$ parameters is equal to 4 for $(\nu_{\underline{m}}^{13}, \underline{m} \in \mathcal{I}_{13})$ plus 4 for $(\nu_{\underline{m}}^{34}, \underline{m} \in \mathcal{I}_{34})$ plus 4 for $(\nu_{\underline{m}}^{24}, \underline{m} \in \mathcal{I}_{24})$ plus 8 for $(\nu_{\underline{m}}^{345}, \underline{m} \in \mathcal{I}_{345})$. We therefore have a total of 20 parameters. The separators are $\{3\}$ and $\{4\}$. We have

$$\{3\} \xrightarrow{o'} \{1, 3\} \text{ and } \{3\} \xrightarrow{o} \{3, 4, 5\}$$

and therefore $N_{\{3\}} = 2$. Similarly

$$\{4\} \xrightarrow{o} \{2, 4\} \text{ and } \{4\} \xrightarrow{o'} \{3, 4, 5\}$$

and therefore $N_{\{4\}} = 2$. According to (32), the dimension of the \mathcal{P} -Dirichlet family is

$$\mathcal{N}_{\mathcal{P}} = 20 - 2 - 2 = 16.$$

For the hyper Dirichlet, the cliques $\{1, 3\}$, $\{3, 4, 5\}$ and $\{2, 4\}$ yield respectively 4, 8 and 4 parameters for a total of 16 while the N_S are the same as in the case of the \mathcal{P} -Dirichlet and therefore, according to (33),

$$\mathcal{N}_{HP} = 16 - 2 - 2 = 12.$$

5.2. Conjugacy and directed strong hyper Markov property. We will now emphasize the properties of the \mathcal{P} -Dirichlet that make it a useful prior for model selection in a restricted class of DAG's. In the following proposition, we state that for any $\mathbf{p} \in \mathcal{P}$, the \mathcal{P} -Dirichlet is strong directed hyper Markov and conjugate. We now recall the definition of the strong directed hyper Markov property. Let $X = (X_1, \dots, X_d)$ be a random variable Markov with respect to a DAG given by a parent function \mathbf{p} , with distribution parameterized by $\theta \in \mathbb{R}^k$ for some k , which itself follows a law \mathcal{L} . Let $\theta_{\mathbf{p}_v}$, $\theta_{v|\mathbf{p}_v}$ and θ_{nd_v} denote respectively the parameters of the marginal distribution of $X_{\mathbf{p}_v}$, the conditional distribution of X_v given $X_{\mathbf{p}_v}$ and the marginal distribution of the non descendants of v . Then the law \mathcal{L} is said to be strong directed hyper Markov if we have the conditional independences

$$\theta_{v|\mathbf{p}_v} \perp \theta_{nd_v} \mid \theta_{\mathbf{p}_v}, \quad v \in V.$$

With this definition, we see that the strong directed hyper Markov property of the \mathcal{P} -Dirichlet follows by construction. We now state and prove that it is a conjugate family.

Theorem 5.2. *Let the conditional distribution of cell counts $\mathbf{N} = (\mathbf{N}(\underline{i}), \underline{i} \in \mathcal{I})$ for $\mathbf{X} = (X_v, v \in V)$ given $\mathbf{p} = (\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ be multinomial $\mathcal{M}_G(\mathbf{p}(\underline{i}), \underline{i} \in \mathcal{I})$ Markov with respect to the decomposable graph G . Let \mathbf{p} follow a \mathcal{P} -Dirichlet distribution with hyper parameters*

$$\nu_{\underline{m}}^A, \underline{m} \in \mathcal{I}_A, A \in \mathfrak{Q} \quad \text{and} \quad \mu_{\underline{n}}^B, \underline{n} \in \mathcal{I}_B, B \in \mathfrak{P}$$

as given in (25) and (26). Then the posterior distribution of \mathbf{p} given $\mathbf{N} = (n(\underline{i}), \underline{i} \in \mathcal{I})$ is \mathcal{P} -Dirichlet with hyper parameters

$$n_{\underline{m}}^A + \nu_{\underline{m}}^A, \underline{m} \in \mathcal{I}_A, A \in \mathfrak{Q} \quad \text{and} \quad n_{\underline{n}}^B + \mu_{\underline{n}}^B, \underline{n} \in \mathcal{I}_B, B \in \mathfrak{P},$$

where $n_{\underline{m}}^A$ is the A -marginal count for $\underline{i}_A = \underline{m}$.

Moreover, for any $\mathbf{p} \in \mathcal{P}$, the \mathcal{P} -Dirichlet is strong hyper Markov.

Proof. The conditional distribution of \mathbf{N} given \mathbf{p} has the density (with respect to the counting measure) which, up to a multiplicative scalar, is equal to $\prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{N(\underline{i})}$. Then, by the generalized Bayes formula for any table $\mathbf{r} = (r(\underline{i}), \underline{i} \in \mathcal{I})$ of nonnegative integers

$$\mathbb{E} \left(\prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r(\underline{i})} \middle| \mathbf{N} = (n(\underline{i}), \underline{i} \in \mathcal{I}) \right) = \frac{\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r(\underline{i}) + n(\underline{i})}}{\mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{n(\underline{i})}}.$$

Applying (25) to the numerator and denominator we see that the right hand side above can be written as

$$\frac{\prod_{A \in \mathfrak{Q}} \prod_{\underline{m} \in \mathcal{I}_A} (\nu_{\underline{m}}^A)^{r_{\underline{m}}^A + n_{\underline{m}}^A}}{\prod_{A \in \mathfrak{Q}} \prod_{\underline{m} \in \mathcal{I}_A} (\nu_{\underline{m}}^A)^{n_{\underline{m}}^A}} \frac{\prod_{B \in \mathfrak{P}} \prod_{\underline{n} \in \mathcal{I}_B} (\mu_{\underline{n}}^B)^{r_{\underline{n}}^B}}{\prod_{B \in \mathfrak{P}} \prod_{\underline{n} \in \mathcal{I}_B} (\mu_{\underline{n}}^B)^{r_{\underline{n}}^B + n_{\underline{n}}^B}}.$$

Note that

$$\frac{(\nu_{\underline{m}}^A)^{r_{\underline{m}}^A + n_{\underline{m}}^A}}{(\nu_{\underline{m}}^A)^{n_{\underline{m}}^A}} = (\nu_{\underline{m}}^A + n_{\underline{m}}^A)^{r_{\underline{m}}^A} \quad \text{and} \quad \frac{(\mu_{\underline{n}}^B)^{r_{\underline{n}}^B}}{(\mu_{\underline{n}}^B)^{r_{\underline{n}}^B + n_{\underline{n}}^B}} = (\mu_{\underline{n}}^B + n_{\underline{n}}^B)^{r_{\underline{n}}^B}.$$

Consequently,

$$E \left(\prod_{\underline{i} \in \mathcal{I}} [\mathbf{p}(\underline{i})]^{r(\underline{i})} \middle| \mathbf{N} \right) = \frac{\prod_{A \in \mathfrak{Q}} \prod_{\underline{m} \in \mathcal{I}_A} (\nu_{\underline{m}}^A + n_{\underline{m}}^A)^{r_{\underline{m}}^A}}{\prod_{B \in \mathfrak{P}} \prod_{\underline{n} \in \mathcal{I}_B} (\mu_{\underline{n}}^B + n_{\underline{n}}^B)^{r_{\underline{n}}^B}}$$

and thus it follows from (25) and the fact that the distribution is uniquely determined by moments that the posterior distribution of \mathbf{p} given the counts $\mathbf{N} = \mathbf{n}$ is \mathcal{P} -Dirichlet with parameters updated by counts. We note that the parameters: $\nu_{\underline{m}}^A + n_{\underline{m}}^A$, $A \in \mathfrak{Q}$, and $\mu_{\underline{n}}^B + n_{\underline{n}}^B$, $B \in \mathfrak{P}$, of the posterior distribution of \mathbf{p} satisfy the constraints of the type (26). Indeed, this is due to the facts that these constraints are linear in the parameters, that the original parameters $\nu_{\underline{m}}^A$ and $\mu_{\underline{n}}^B$ satisfy such constraints by assumption and that such constraints are also trivially satisfied by the counts $n_{\underline{m}}^A$ and $n_{\underline{n}}^B$. This proves that the \mathcal{P} -Dirichlet forms a conjugate family of distribution.

The directed strong hyper Markov property of the \mathcal{P} -Dirichlet holds true for every $\mathbf{p} \in \mathcal{P}$ because of the independences (see Def. 3.1) contained in its construction. \square

5.3. Applications. Even in a small example such as Example 5.1, we see that, by going from the hyper Dirichlet to the \mathcal{P} -Dirichlet, we have substantially increased the number of parameters and that the parameters $\nu_{\underline{m}}^{345}$, $\underline{m} \in \mathcal{I}_{345}$ are no longer constrained to add up to μ as they would be in the hyper Dirichlet. We therefore have more flexibility for the choice of the $(\nu_{\underline{m}}^{345}, m \in \mathcal{I}_{345})$ than we would have in the hyper Dirichlet.

Let us assume that a preliminary studies has determined that the set of conditional independences between these variables is represented by the decomposable graph G in our example. We now want to find the DAG model that best fits the data. But let us assume also that we are told by an expert that vertices 3 and 4 must be parents of 5. In our DAG model search, it then does not make sense to include DAG's that contain directed edges going from 5 to 4 or from 5 to 3. The hyper Dirichlet is a prior on the $(p(\underline{i}), \underline{i} \in \mathcal{I})$ that includes all possible DAG and therefore the right thing to do, in that case, is to take the \mathcal{P} -Dirichlet as a prior in order to exclude the possibilities of having arrow going from 5 to 4 or 3. Since, as pointed out above, we have more flexibility in the choice of the $\nu_{\underline{m}}^{345}$, $m \in \mathcal{I}_{345}$, we can choose them to put more prior weight on the edges $4 \rightarrow 5$ and $3 \rightarrow 5$ if we wish.

Such situations are discussed, for example, in Angelopoulos and Cussens (2008). The focus of that paper is structure learning and they advocate the use of independent Dirichlet priors on the $\mathbf{p}_{m|\underline{k}}^{v|p_v}$, which is precisely what the \mathcal{P} -Dirichlet does for a selected number of orders. There is, of course, the restriction that the underlying graph is decomposable but searches can often be started in the space of decomposable graphical models. Situations where one would typically want to use the \mathcal{P} -Dirichlet would be ones where conditional independence according to a decomposable graphs are imposed by an expert and where, moreover, we are told that a variable must be a parent of another or a variable must be a root node or a variable must be a leaf node or any other situation where a pattern is imposed for some of the directed edges.

6. CHARACTERIZATIONS BY LOCAL AND GLOBAL INDEPENDENCE

6.1. The \mathcal{P} -Dirichlet and the hyper-Dirichlet. We now briefly recall the definition of local and global independence

Definition 6.1. Let \mathbf{p} be a vector of random probabilities associated with the graph G . We say that local parameter independence holds for \mathbf{p} with respect to a DAG with a parent function \mathbf{p} if for any $v \in V$ the random vectors

$$(\mathbb{P}_{\mathbf{p}}(X_v = l | X_{\mathbf{p}_v} = \underline{k}), l \in \mathcal{I}_v), \quad \underline{k} \in \mathcal{I}_{\mathbf{p}_v},$$

are independent (non-degenerate) and we say that global parameter independence holds for \mathbf{p} if the random vectors

$$((\mathbb{P}_{\mathbf{p}}(X_v = l | X_{\mathbf{p}_v} = \underline{k}), l \in \mathcal{I}_v), \quad \underline{k} \in \mathcal{I}_{\mathbf{p}_v}), \quad v \in V$$

are independent (non-degenerate).

It appears that for such families of DAGs local and global independence of parameters characterizes the \mathcal{P} -Dirichlet distribution. Actually, we do not need to distinguish between the two properties. Therefore instead we combine them into one property of *parameter independence*.

Definition 6.2. Let \mathbf{p} be a vector of random probabilities associated with the graph G . We say that parameter independence holds for \mathbf{p} with respect to a DAG with a parent function \mathbf{p} if both local and global independence hold for \mathbf{p} or equivalently if the random vectors

$$(\mathbb{P}_{\mathbf{p}}(X_v = l | X_{\mathbf{p}_v} = \underline{k}), l \in \mathcal{I}_v), \quad \underline{k} \in \mathcal{I}_{\mathbf{p}_v}, \quad v \in V$$

are independent (non-degenerate).

Analogously, we say that parameter independence holds for \mathbf{p} with respect to a family \mathcal{P} of DAGs if it holds for any DAG from \mathcal{P} .

We immediately note that under the condition of parameter independence with respect to a DAG having a parent function \mathbf{p} for any d -way table $\underline{r} = (r_{\underline{i}}, \underline{i} \in \mathcal{I})$ of non-negative integers we have

$$(35) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i})]^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}_v}} \mathbb{E} \prod_{l \in \mathcal{I}_v} [\mathbb{P}_{\mathbf{p}}(X_v = l | X_{\mathbf{p}_v} = \underline{k})]^{r_{\underline{k}, l}^{\mathbf{q}_v}},$$

where

$$(36) \quad r_{\underline{k}, l}^{\mathbf{q}_v} = \sum_{\underline{i} \in \mathcal{I}: \underline{i}_{\mathbf{q}_v} = (\underline{k}, l)} r_{\underline{i}}.$$

This property of parameter independence was crucial for the characterization of the Dirichlet distribution (for a complete graph) given in Geiger and Heckerman (1997). We will extend this characterization to the \mathcal{P} -Dirichlet distribution for various families \mathcal{P} of DAGs including those yielding the hyper-Dirichlet law. Such families are described by what we call a separating property.

Definition 6.3. A family \mathcal{P} of DAGs with structure G is called separating if

$$(37) \quad \forall v \in V \quad \exists \mathbf{p}, \mathbf{p}' \in \mathcal{P} \quad \text{such that} \quad \mathbf{p}_v \neq \mathbf{p}'_v,$$

In our main result we will show that for separating families of DAGs parameter independence characterizes the \mathcal{P} -Dirichlet distribution.

Theorem 6.1. Let \mathbf{p} be a vector of random probabilities associated with the graph G . Let \mathcal{P} be a separating family of moral DAGs for a decomposable graph $G = (V, E)$.

If parameter independence for \mathbf{p} with respect to \mathcal{P} holds then \mathbf{p} has a \mathcal{P} -Dirichlet distribution.

The proof is given in the Appendix.

The hyper-Dirichlet distribution is similarly characterized by parameter independence with respect to a \mathcal{P} which is rich enough. We state this result more precisely in the theorem below which is an immediate consequence of Theorems 6.1 and 4.6.

Theorem 6.2. Let \mathbf{p} be a vector of random probabilities associated with a decomposable graph G . Let \mathcal{P} , a separating family of moral DAGs with skeleton G , satisfies (30) and (31).

If parameter independence for \mathbf{p} with respect to \mathcal{P} holds then \mathbf{p} has a hyper-Dirichlet distribution.

6.2. Special cases.

6.2.1. *The chain and the hyper Dirichlet distribution.* Let $G = (V, E)$ be a chain with vertices $V = \{1, \dots, d\}$ and edges $E = \{\{i, i+1\}, i = 1, \dots, d-1\}$. Then $\mathcal{C} = E$ and $\mathcal{S} = \{\{2\}, \dots, \{d-1\}\}$.

Consider the family $\mathcal{P} = \{\mathbf{p}, \mathbf{p}'\}$, where

$$\mathbf{p}_1 = \emptyset \quad \text{and} \quad \mathbf{p}_i = i-1, \quad i = 2, \dots, d,$$

and

$$\mathbf{p}'_d = \emptyset \quad \text{and} \quad \mathbf{p}'_i = i+1, \quad i = 1, \dots, d-1.$$

Note that \mathcal{P} is separating. Note also that

$$\mathfrak{R}_{\mathbf{p}} = \mathbf{p}(V) \setminus \mathcal{S} = \{\{1\}\} \quad \text{and} \quad \mathfrak{R}_{\mathbf{p}'} = \mathbf{p}'(V) \setminus \mathcal{S} = \{\{d\}\}.$$

Therefore $\mathfrak{R} = \emptyset$ and (30) holds. Moreover, condition (31) is clearly satisfied. Thus from Theorem 6.2 we conclude the following result.

Corollary 6.3. *Assume that the random vectors*

$$(38) \quad (\mathbb{P}_{\mathbf{p}}(X_j = l | X_{j-1} = k), l \in \mathcal{I}_j), \quad k \in \mathcal{I}_{j-1}, \quad j = 1, 2, \dots, d$$

are jointly independent.

Assume also that the random vectors

$$(39) \quad (\mathbb{P}_{\mathbf{p}}(X_j = l | X_{j+1} = m), l \in \mathcal{I}_j), \quad m \in \mathcal{I}_{j+1}, \quad j = 1, 2, \dots, d$$

are jointly independent.

Then \mathbf{p} has a hyper Dirichlet distribution with respect to G .

In the assumptions above we used the convention that $X_0 = X_{d+1} = 0$ and $\mathcal{I}_0 = \mathcal{I}_{d+1} = \{0\}$.

Note that the family \mathcal{P} that we defined for the chain is the unique minimal separating family. That is any other family of DAGs is either non-separating or it contains \mathcal{P} as a proper subset.

Note also that for the two-chain (that is when $d = 2$) we obtain the characterization of the classical Dirichlet distribution given in Th. 2 of Bobecka and Wesolowski (2009). At the same time we extend the characterization given in Geiger and Heckerman (1997), Th. 2, where additionally to parameter independences as above it was assumed that densities exist and are sufficiently regular. Some of the regularity assumptions were considerably weakened in J  rai (1998). More recently the entire Ch. 23 of the monograph J  rai (2005) was devoted to this issue.

6.2.2. *The tree and the hyper Dirichlet distribution.* Let $G = (V, E)$ be a tree. As in the case of the chain the set of cliques \mathcal{C} is equal to E and $\mathcal{S} = \{\{v\} : v \in V \setminus L\}$, where $L \subset V$ is the set of leaves, that is, those vertices which belong to exactly one edge. Any DAG can be uniquely defined by choosing a vertex v such that $\mathbf{p}_v = \emptyset$. We will denote this DAG by \mathcal{G}_v . Note that for any such \mathcal{G}_v , for any $w \in V \setminus \{v\}$, the set \mathbf{p}_w contains exactly one element.

Consider the family $\mathcal{P} = \{\mathcal{G}_v, v \in L\}$. Note that each vertex on the unique chain connecting v and w in L has different parents in \mathcal{G}_v and \mathcal{G}_w and therefore \mathcal{P} is a separating family. Since any separator consists of only one vertex and any clique of only two vertices condition (31) follows from the same observation. Since $\mathbf{p}_v(V) = \mathcal{S} \cup \{v\}$ it follows that $\mathfrak{R}_{\mathbf{p}_v} = \{\{v\}\}$. Consequently, (30) is satisfied. From Theorem 6.2 we have the following result for trees.

Corollary 6.4. *Assume that for every leaf $v \in L$ the condition of parameter independence with respect to \mathcal{G}_v holds.*

Then $(\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i}), \underline{i} \in \mathcal{I})$ follows the hyper Dirichlet distribution with respect to G .

6.2.3. The complete graph and the classical Dirichlet distribution. Consider a complete graph $G = (V, E)$, for which we, of course, have $\mathcal{C} = \{V\}$ and $\mathcal{S} = \emptyset$. Consider two DAGs with parent functions \mathbf{p} and \mathbf{p}' . Let v_1, \dots, v_d be the numbering corresponding to \mathbf{p} and v'_1, \dots, v'_d that corresponding to \mathbf{p}' . Let us assume moreover that

$$(40) \quad \forall j = 1, \dots, d \quad \mathbf{p}_{v_j} \neq \mathbf{p}'_{v'_j}.$$

We claim that the family $\mathcal{P} = \{\mathbf{p}, \mathbf{p}'\}$ is separating. Since the cardinality of \mathbf{p}_{v_j} is equal to $j - 1$, $\mathbf{p}_{v_j} = \mathbf{p}_{v'_k}$ implies $j = k$. But this is forbidden for $j = 1, \dots, d - 1$ by condition (40). Therefore the family is separating. Similarly, $\mathfrak{R}_{\mathbf{p}} = \mathbf{p}(V) = \{\mathbf{p}_{v_j}, j = 2, \dots, d\}$ and $\mathfrak{R}_{\mathbf{p}'} = \mathbf{p}'(V) = \{\mathbf{p}'_{v'_j}, j = 2, \dots, d\}$ cannot have a common element by the cardinality argument combined with (40). Thus (30) is satisfied. We therefore have the following result.

Corollary 6.5. *If for a complete graph $G = (V, E)$ parameter independence holds for any two DAGs satisfying (40) then $(\mathbb{P}_{\mathbf{p}}(\underline{X} = \underline{i}), \underline{i} \in \mathcal{I})$ has a classical Dirichlet distribution.*

For $d = 2$ this is the case of the chain discussed earlier in this section. In Heckerman, Geiger and Chickering (1995), Th. 7 (see also Geiger and Heckerman (1997), Th.3) the authors considered parameter independence for a very special family of two DAGs with parent functions \mathbf{p} and \mathbf{p}' defined as follows

$$\mathbf{p}_1 = \emptyset, \quad \mathbf{p}_i = \{1, \dots, i - 1\}, \quad i = 2, \dots, d,$$

and

$$\mathbf{p}'_i = \{d\} \cup \mathbf{p}_i, \quad i = 1, \dots, d - 1, \quad \mathbf{p}'_d = \emptyset.$$

Such a choice of family of DAGs was important for their proof in the case $d > 2$, since then they could easily reduce the problem to the case $d = 2$. As mentioned earlier, this case was settled in Geiger and Heckerman (1997) under some smoothness assumptions on the density. Clearly, \mathbf{p} and \mathbf{p}' above satisfy (40) and their result follows by Corollary 6.5 without assuming that the density exists.

Bobecka and Wesolowski (2009) have exactly the same result as Corollary 6.5 for $d = 2$ only. Since they were interested in, so called, neutralities with respect to partitions, for their characterization of the Dirichlet for $d > 2$, it was natural to assume parameter independences for the parameters of the marginal probability of one variable X_l and those of the conditional probabilities of $\mathbf{X}_{V \setminus \{l\}}$ given X_l . They assumed that for any $l = 1, \dots, d$,

$$(\mathbb{P}_{\mathbf{p}}(X_l = i), i \in \mathcal{I}_l), \quad (\mathbb{P}_{\mathbf{p}}(\mathbf{X}_{V \setminus \{l\}} = \underline{i}_{V \setminus \{l\}} | X_l = i), \underline{i}_{V \setminus \{l\}} \in \mathcal{I}_{V \setminus \{l\}}), \quad i \in \mathcal{I}_l,$$

are independent. The approach was via moments and no density assumptions were required. A related result based on purely Bayesian argument was obtained in Ramamoorthi and Sangalli (2007). For related characterizations of the classical Dirichlet, see, e.g., Darroch and Ratcliff (1971), Fabius (1973), James and Mosimann (1980), Bobecka and Wesolowski (2007), Chang et al. (2010), Sakowicz and Wesolowski (2014) and the monograph Ng et al. (2011), Ch. 2.6.

7. CONCLUSION

This paper makes two contributions. The first one is practical and the second theoretical. Let us consider a given set of variables and a given set of conditional independences between these variables which can be represented graphically by means of a moral DAG with skeleton G , decomposable. Assume that \mathcal{P} is the collection of a certain number of these moral DAG's and that we want to perform model selection in that class of DAG's. Such a situation may arise when we are given a set of conditional independences and we are also given the direction of certain edges. In that case, we want to put a zero prior probability on any DAG that does not follow these requirements. Our new \mathcal{P} -Dirichlet does just that and it has the advantage of increased flexibility in the choice of hyper parameters (due to the restricted number of DAG's in \mathcal{P}). As we showed, the \mathcal{P} -Dirichlet forms a conjugate family of prior distribution with the strong hyper Markov property.

Our theoretical contribution is a characterization based on local and global parameter independence of this new family of distributions and in particular of the hyper Dirichlet and the classical Dirichlet without the assumption of the existence of the density.

We would also like to emphasize that in the development of our new prior distribution, we have introduced new objects such as the sets \mathfrak{P} and \mathfrak{Q} generalizing the notion of cliques and separators in a decomposable graph G . We have also shed light on the choice of two DAG's on a complete subset used by Geiger and Heckerman (1997) in their characterization by emphasizing that these two DAG's form a separating family when G is complete and that this particular choice of two DAG's is only one of many possible choices.

Acknowledgement. H. Massam gratefully acknowledges support from NSERC Discovery Grant No A8947. J. Wesolowski was partially supported by NCN grant 2012/05/B/ST1/00554.

References

- (1) ANGELOPOULOS, N. AND CUSSENS, J., Bayesian learning of Bayesian networks with informative priors. *Annals of Mathematics and Artificial Intelligence*, **54**(1-3) (2008), 53-98.
- (2) ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. AND TRIGGS, C. M. , On the relation between conditional independence models determined by finite distributive lattices and directed acyclic graphs. *J. Statist. Plann. Infer.* **48** (1995), 25-46.
- (3) BOBECKA, K., WESOŁOWSKI, J., The Dirichlet distribution and process through neutralities. *J. Theor. Probab.* **20** (2007), 295-308.
- (4) BOBECKA, K., WESOŁOWSKI, J., Moments approach to characterizations of Dirichlet tables through neutralities. *Publ. Math. Debrecen* **74/3-4** (2009), 321-339.
- (5) CHANG, W.Y., GUPTA, R.D., RICHARDS, D.St.P., Structural properties of the generalized Dirichlet distributions. In: *Algebraic Methods in Statistics and Probability, II* (M.A.G. Viana, H.P. Wynn, eds), *Contemp. Math.* **516** (2010), 109-124.
- (6) DARROCH, J., RATCLIFF, D., A characterization of the Dirichlet distribution. *J. Amer. Statist. Assoc.* **66** (1971), 641-643.
- (7) DAWID, P., LAURITZEN, S.L., Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** (1993), 1275-1317.

- (8) DIACONIS, P., YLVISAKER, D., Conjugate priors for exponential families. *Ann. Statist.* **7** (1979), 269-281.
- (9) FABUS, J., Two characterizations of the Dirichlet distribution. *Ann. Statist.* **1** (1973), 583-587.
- (10) GEIGER, D., HECKERMAN, D., A characterization of the Dirichlet distribution through global and local parameter independence. *Ann. Statist.* **25** (1997), 1344-1369.
- (11) HECKERMAN, D. GEIGER, D., CHICKERING, D.M., Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learn.* **20** (1995), 197-243.
- (12) JAMES, I.R., MOSIMANN, J.E., A new characterization of the Dirichlet distribution through neutrality. *Ann. Statist.* **8(1)** (1980), 183-189.
- (13) LAURITZEN, S.L., *Graphical Models*. Oxford Univ. Press, Oxford, 1996.
- (14) JÁRAI, A., Regularity property of the functional equation of the Dirichlet distribution. *Aeq. Math.* **56** (1998), 37-46.
- (15) JÁRAI, A., *Regularity Properties of Functional Equations in Several Variables*. Springer, New York, 2005.
- (16) NG, K.W., TIAN, G.-L., TANG, M.-L., *Dirichlet and Related Distributions, Theory, Methods and Applications*, Wiley, New York, 2011.
- (17) RAMAMOORTHI, R.V., SANGALLI, L.M., On a characterization of the Dirichlet distribution. In: *Bayesian Statistics and Its Applications*, (S. K. Upadhyay, U. Singh, D. K. Dey, eds.), Anamaya Publishers, New Delhi, 2007, 385-397.
- (18) SAKOWICZ, A., WESOŁOWSKI, J., Dirichlet distribution through neutralities with respect to two partitions. *J. Multivar. Anal.* **129** (2014), 1-15.
- (19) SPIEGELHALTER, D., LAURITZEN, S., Sequential updating of conditional probabilities on directed graphical structures. *Networks*, **20**, 579-605, 1990.

8. APPENDIX

8.1. Proof of Theorem 4.3.

Proof. Assume that $(\mathbf{p}(\underline{i}), i \in \mathcal{I})$ is \mathcal{P} -Dirichlet distributed. That is representation (15) holds and for any $\mathbf{p} \in \mathcal{P}$, the random vectors $(\mathbf{p}_{m|\underline{k}}^{v|\mathbf{p}_v})_{m \in \mathcal{I}_v}$ follow a Dirichlet distribution $\text{Dir}(\alpha_{m|\underline{k}}^{v|\mathbf{p}_v}, m \in \mathcal{I}_v)$, $\underline{k} \in \mathcal{I}_{\mathbf{p}_v}$, $v \in V$ and are independent. That is, we have to identify the parameters $\alpha_{m|\underline{k}}^{v|\mathbf{p}_v}$, $m \in \mathcal{I}_v$, $\underline{k} \in \mathcal{I}_{\mathbf{p}_v}$, $v \in V$, such that (see (16) and (17))

$$(41) \quad \mathbb{E} \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}_v}} \prod_{m \in \mathcal{I}_v} \left(\mathbf{p}_{m|\underline{k}}^{v|\mathbf{p}_v} \right)^{r_{(\underline{k}, m)}^{qv}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}_v}} \frac{\prod_{m \in \mathcal{I}_v} (\alpha_{m|\underline{k}}^{v|\mathbf{p}_v})^{r_{(\underline{k}, m)}^{qv}}}{(\tilde{\alpha}_{\underline{k}}^{\mathbf{p}_v})^{r_{\underline{k}}^{\mathbf{p}_v}}},$$

where

$$(42) \quad \tilde{\alpha}_{\underline{k}}^{\mathbf{p}_v} = \sum_{m \in \mathcal{I}_v} \alpha_{m|\underline{k}}^{v|\mathbf{p}_v} \quad \forall \underline{k} \in \mathcal{I}_{\mathbf{p}_v},$$

is equal to the right-hand side of (25) with consistency conditions given in (26).

Note that the result in the opposite direction, that is the fact that the formula (25) for moments, together with the constraints (26), implies \mathcal{P} -Dirichlet distribution for $(\mathbf{p}(\underline{i}), i \in \mathcal{I})$, follows then immediately from the property that the distribution is uniquely determined by moments.

Consider an arbitrary collection $O_{\mathcal{P}}$ of \mathbf{p} -perfect orders, $\mathbf{p} \in \mathcal{P}$. Fix an arbitrary $\mathbf{p} \in \mathcal{P}$ and consider a \mathbf{p} -perfect order $o \in O_{\mathcal{P}}$. We will now relate the sets Q_i^C , $i = 0, 1, \dots, j_C$, $C \in \mathcal{C}$, to the numbering of vertices imposed by o as given in (3). Clearly $C = C_l$ for some $l \in \{1, \dots, K\}$. For ease of notation, we suppress the subscript l in the remainder of this proof. We define $j(l, i)$, $i \in \{1, \dots, j_C - 1\}$ to be the index of the vertex which satisfies

$$(43) \quad Q_i^C = \mathbf{q}_{v_{l,j(l,i)}} = \mathbf{p}_{v_{l,j(l,i)+1}}.$$

We also note that

$$(44) \quad C = Q_0^C = \mathbf{q}_{v_{l,C_l}} \quad \text{and} \quad S = S_l = Q_{j_C}^C = \mathbf{p}_{v_{l,s_l+1}}.$$

We will now define the $\alpha_{m|\underline{k}}^{v|\mathbf{p}_v}$'s in terms of the ν^A 's. For any $l \in \{1, \dots, K\}$, if $v = v_{l,j(l,i)}$, set

$$(45) \quad \alpha_{m|\underline{k}}^{v|\mathbf{p}_v} := \nu_{(\underline{k},m)}^{Q_i^C} \quad \forall (\underline{k}, m) \in \mathcal{I}_{Q_i^C}.$$

For any $l \in \{1, \dots, K\}$, if $v = v_{l,j}$ and $j \neq j(l, i)$ for any i , we define

$$(46) \quad i_j = \min\{i : \mathbf{q}_v \subset Q_i^{C_l}\}$$

and then set

$$(47) \quad \alpha_{m|\underline{k}}^{v|\mathbf{p}_v} := \sum_{\underline{n} \in \mathcal{I}_{Q_{i_j}^C} \setminus \mathbf{q}_v} \nu_{(\underline{n},m,k)}^{Q_{i_j}^C} \quad \forall (\underline{k}, m) \in \mathcal{I}_{\mathbf{q}_v}.$$

We will now show that, if $v_{l,j}$ is such that $j \neq j(l, i)$ for any $i \in \{0, \dots, j_C - 1\}$, then

$$(48) \quad \alpha_{m|\underline{k}}^{v_{l,j}|\mathbf{p}_{v_{l,j}}} = \tilde{\alpha}_{(\underline{k},m)}^{\mathbf{p}_{v_{l,j}+1}} \quad \forall (\underline{k}, m) \in \mathcal{I}_{\mathbf{p}_{v_{l,j}+1}}.$$

Consider first the case when $j(l, i_j) = j + 1$. By (47) we have

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathbf{p}_{v_{l,j}}} = \sum_{n \in \mathcal{I}_{v_{l,j}+1}} \nu_{(n,\underline{k},m)}^{Q_{i_j}^C}$$

Since $Q_{i_j}^C = \mathbf{q}_{v_{l,j}+1}$ from the above equality and (45) we get

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathbf{p}_{v_{l,j}}} = \sum_{\underline{n} \in \mathcal{I}_{v_{l,j}+1}} \alpha_{n|(\underline{k},m)}^{v_{l,j}+1|\mathbf{p}_{v_{l,j}+1}}.$$

Thus (48) follows from (42).

Second, consider the case $j(l, i_j) > j + 1$. By (47) we have

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathbf{p}_{v_{l,j}}} = \sum_{\underline{n} \in \mathcal{I}_{Q_{i_j}^C} \setminus \mathbf{q}_{v_{l,j}}} \nu_{(\underline{n},\underline{k},m)}^{Q_{i_j}^C} = \sum_{n_1 \in \mathcal{I}_{v_{l,j}+1}} \sum_{\underline{n}_2 \in \mathcal{I}_{Q_{i_j}^C} \setminus \mathbf{q}_{v_{l,j}+1}} \nu_{(n_1,\underline{n}_2,\underline{k},m)}^{Q_{i_j}^C},$$

where the second equality follows from the fact that $\mathbf{q}_{v_{l,j}+1} = \mathbf{q}_{v_{l,j}} \cup \{v_{l,j}+1\}$. Applying (47) to the inner sum we obtain

$$\alpha_{m|\underline{k}}^{v_{l,j}|\mathbf{p}_{v_{l,j}}} = \sum_{n_1 \in \mathcal{I}_{v_{l,j}+1}} \alpha_{n_1|(\underline{k},m)}^{v_{l,j}+1|\mathbf{p}_{v_{l,j}+1}}.$$

Thus (48) follows from (42).

Due to (48) we have cancelations in the right-hand side of (41) and the only terms left are:

- in the numerator: $\alpha_{m|\underline{k}}^{v|\mathbf{p}_v} = \nu_{(\underline{k},m)}^{Q_i^C}$ for $v = v_{l,j(l,i)}$, where $i \in \{0, \dots, j_C - 1\}$.
- in the denominator: $\tilde{\alpha}_{\underline{k}}^{\mathbf{p}_v}$ for $v = v_{l,j(l,i)+1}$ where $i \in \{1, \dots, j_C\}$.

In particular, $j(l, j_C) = s_l$ in general and for $l = 1$, $s_l = 0$ so that, in the denominator, we have parameters indexed by $\mathbf{p}_{v_{l,s_l+1}} = S_l$ and $\mathbf{p}_{v_{1,1}} = \emptyset$.

To complete the proof, that is to show that the right-hand side of (41) is equal to the right-hand side of (25), it remains to show that for any $l \in \{1, \dots, K\}$ and $\underline{i} \in \{1, \dots, j_C\}$, we have

$$(49) \quad \mu_{\underline{k}}^{Q_i^C} = \tilde{\alpha}_{\underline{k}}^{\mathbf{p}_{v_{l,j(l,i)+1}}} \quad \forall \underline{k} \in \mathcal{I}_{Q_i^C}.$$

Note that

- (i) either $j(l, i) + 1 = j(l, i - 1)$, that is $\mathbf{q}_{v_{l,j(l,i)+1}} = Q_{i-1}^C$,
- (ii) or $j(l, i) + 1$ is not of the form $j(l, \tilde{i})$ for some $\tilde{i} \in \{i + 1, \dots, j_C\}$ (observe - see (46) - that in this case we have $i_{j(l,i)+1} = i - 1$).

Let's consider case (i) first. Using (42) and then (45) for all $\underline{k} \in \mathcal{I}_{Q_i^C}$ (note that $Q_i^C = \mathbf{p}_{v_{l,j(l,i)+1}}$) we obtain

$$\tilde{\alpha}_{\underline{k}}^{\mathbf{p}_{v_{l,j(l,i)+1}}} = \sum_{m \in \mathcal{I}_{v_{l,j(l,i)+1}}} \alpha_{m|\underline{k}}^{v_{l,j(l,i)+1}|\mathbf{p}_{v_{l,j(l,i)+1}}} = \sum_{m \in \mathcal{I}_{v_{l,j(l,i)+1}}} \nu_{(\underline{k},m)}^{Q_{i-1}^C}.$$

Since $\{v_{l,j(l,i)+1}\} = \mathbf{q}_{v_{l,j(l,i)+1}} \setminus \mathbf{q}_{v_{l,j(l,i)}}$ due to (26) we obtain (49).

For case (ii), we use again (42) and then (47) to arrive at

$$\tilde{\alpha}_{\underline{k}}^{\mathbf{p}_{v_{l,j(l,i)+1}}} = \sum_{m \in \mathcal{I}_{v_{l,j(l,i)+1}}} \sum_{\underline{n} \in \mathcal{I}_{Q_{i-1}^C} \setminus \mathbf{q}_{v_{l,j(l,i)+1}}} \nu_{(m,\underline{n},\underline{k})}^{Q_{i-1}^C} = \sum_{(m,\underline{n}) \in \mathcal{I}_{Q_{i-1}^C} \setminus \mathbf{q}_{v_{l,j(l,i)}}} \nu_{(m,\underline{n},\underline{k})}^{Q_{i-1}^C},$$

where the last equation follows from the fact that $\mathbf{q}_{v_{l,j(l,i)+1}} = \mathbf{q}_{v_{l,j(l,i)}} \cup \{v_{l,j(l,i)+1}\}$. Moreover $\mathbf{q}_{v_{l,j(l,i)}} = Q_i^C$, therefore (49) follows now from assumption (26). \square

8.2. Proof of Theorem 6.1.

Proof. Note that according to Def. 3.2 it suffices to show that the formula for moments as given in (41) and (42) holds.

From (35) it follows that for any DAG from \mathcal{P} with parent function \mathbf{p} and any \underline{r}

$$(50) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i})]^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}_v}} f_{\underline{k}}^{\mathbf{p}_v} \left(r_{\underline{k},l}^{q_v}, l \in \mathcal{I}_v \right),$$

where for any $v \in V$ and any $\underline{k} \in \mathcal{I}_{\mathbf{p}_v}$

$$f_{\underline{k}}^{\mathbf{p}_v}(z_l, l \in \mathcal{I}_v) = \mathbb{E} \prod_{l \in \mathcal{I}_v} [\mathbb{P}_{\mathbf{p}}(X_v = l | \mathbf{X}_{\mathbf{p}_v} = \underline{k})]^{z_l}, \quad z_l \in \{0, 1, \dots\}, l \in \mathcal{I}_v.$$

In order to prove Theorem 6.1 we identify the functions $f_{\underline{k}}^{\mathbf{p}_v}$, $\underline{k} \in \mathcal{I}_{\mathbf{p}_v}$, $v \in V$. Our method relies on identification of the general form of the functions $f_{\underline{k}}^{\mathbf{p}_v}$, which will appear to be a ratio of products of gamma functions as in the formula for the moments of the \mathcal{P} -Dirichlet distribution. Our main tool is equation (62) of Lemma 8.1. The proof is divided into two parts, (a) and (b). In part (a) we transform the moment equation (50) into a the seemingly cumbersome but useful (54). In part (b) through a judicious choice of sparse \underline{r} 's in (54) we will obtain the general form of $f_{\underline{k}}^{\mathbf{p}_v}$'s using Lemma 8.1.

(a) We first aim for the simplified functional equation (54). Fix an arbitrary $\underline{\tau} = (\tau_v \in \mathcal{I}_v, v \in V)$ and consider an d -way table $\underline{\epsilon} = (\epsilon_{\underline{i}}, \underline{i} \in \mathcal{I})$ such that

$$\epsilon_{\underline{i}} = \begin{cases} 1, & \text{if } \underline{i} = \underline{\tau}, \\ 0, & \text{otherwise.} \end{cases}$$

Changing $\underline{\tau}$ into $\underline{\tau} + \underline{\epsilon}$ in (50) we get

$$(51) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i})]^{r_{\underline{i}} + \epsilon_{\underline{i}}} = \prod_{v \in V} [f_{\tau_{\mathbf{p}v}}^{\mathbf{p}v} (r_{\tau_{\mathbf{p}v}}^{\mathbf{q}v} + 1, r_{\tau_{\mathbf{p}v}, l}^{\mathbf{q}v}, l \in \mathcal{I}_v \setminus \{\tau_v\})] \\ \times \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}v} \setminus \{\tau_{\mathbf{p}v}\}} f_{\underline{k}}^{\mathbf{p}v} (r_{\underline{k}, l}^{\mathbf{q}v}, l \in \mathcal{I}_v)]$$

We will now obtain an equation of the type (62) by equating the right-hand side of (51) for different \mathbf{p} 's from \mathcal{P} . Fix a DAG in \mathcal{P} , that is a $\mathbf{p} \in \mathcal{P}$, and fix a vertex $v \in V$. Then, by separation property (37) there exists another DAG in \mathcal{P} with parent function \mathbf{p}' such that $\mathbf{p}'_v \neq \mathbf{p}_v$. For each of \mathbf{p} and \mathbf{p}' the right-hand side of (51) is split into three parts: the first (first line) concerns v , the second (second line) \mathbf{c}_v and the third (third line) the remainder of V . Thus we obtain

$$(52) \quad \begin{aligned} & f_{\tau_{\mathbf{p}v}}^{\mathbf{p}v} (r_{\tau_{\mathbf{p}v}}^{\mathbf{q}v} + 1, r_{\tau_{\mathbf{p}v}, l}^{\mathbf{q}v}, l \in \mathcal{I}_v \setminus \{\tau_v\}) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}v} \setminus \{\tau_{\mathbf{p}v}\}} f_{\underline{k}}^{\mathbf{p}v} (r_{\underline{k}, l}^{\mathbf{q}v}, l \in \mathcal{I}_v) \\ & \prod_{w \in \mathbf{c}_v} f_{\tau_{\mathbf{p}w}}^{\mathbf{p}w} (r_{\tau_{\mathbf{p}w}}^{\mathbf{q}w} + 1, r_{\tau_{\mathbf{p}w}, l}^{\mathbf{q}w}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}w} \setminus \{\tau_{\mathbf{p}w}\}} f_{\underline{k}}^{\mathbf{p}w} (r_{\underline{k}, l}^{\mathbf{q}w}, l \in \mathcal{I}_w) \\ & \prod_{w \notin \mathbf{c}_v \cup \{v\}} f_{\tau_{\mathbf{p}w}}^{\mathbf{p}w} (r_{\tau_{\mathbf{p}w}}^{\mathbf{q}w} + 1, r_{\tau_{\mathbf{p}w}, l}^{\mathbf{q}w}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}w} \setminus \{\tau_{\mathbf{p}w}\}} f_{\underline{k}}^{\mathbf{p}w} (r_{\underline{k}, l}^{\mathbf{q}w}, l \in \mathcal{I}_w) \\ & = f_{\tau_{\mathbf{p}'v}}^{\mathbf{p}'v} (r_{\tau_{\mathbf{p}'v}}^{\mathbf{q}'v} + 1, r_{\tau_{\mathbf{p}'v}, l}^{\mathbf{q}'v}, l \in \mathcal{I}_v \setminus \{\tau_v\}) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}'v} \setminus \{\tau_{\mathbf{p}'v}\}} f_{\underline{k}}^{\mathbf{p}'v} (r_{\underline{k}, l}^{\mathbf{q}'v}, l \in \mathcal{I}_v) \\ & \prod_{w \in \mathbf{c}'_v} f_{\tau_{\mathbf{p}'w}}^{\mathbf{p}'w} (r_{\tau_{\mathbf{p}'w}}^{\mathbf{q}'w} + 1, r_{\tau_{\mathbf{p}'w}, l}^{\mathbf{q}'w}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}'w} \setminus \{\tau_{\mathbf{p}'w}\}} f_{\underline{k}}^{\mathbf{p}'w} (r_{\underline{k}, l}^{\mathbf{q}'w}, l \in \mathcal{I}_w) \\ & \prod_{w \notin \mathbf{c}'_v \cup \{v\}} f_{\tau_{\mathbf{p}'w}}^{\mathbf{p}'w} (r_{\tau_{\mathbf{p}'w}}^{\mathbf{q}'w} + 1, r_{\tau_{\mathbf{p}'w}, l}^{\mathbf{q}'w}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}'w} \setminus \{\tau_{\mathbf{p}'w}\}} f_{\underline{k}}^{\mathbf{p}'w} (r_{\underline{k}, l}^{\mathbf{q}'w}, l \in \mathcal{I}_w) \end{aligned}$$

We now write this equation above for two distinct values first for $\tau_v = \rho$ and then for $\tau_v = \sigma$ in \mathcal{I}_v , while keeping τ_k the same for all $k \neq v$. We obtain two equations, say E_ρ and E_σ and we then write the identity

$$(53) \quad \frac{\text{lhs}(E_\rho)}{\text{lhs}(E_\sigma)} = \frac{\text{rhs}(E_\rho)}{\text{rhs}(E_\sigma)}.$$

Many simplifications occur (see part 8.4 of Appendix) and we arrive at

$$\begin{aligned}
 & \frac{f_{\tau_{\mathbf{p}v}}^{\mathbf{p}v} \left(r_{\tau_{\mathbf{p}v}, \rho}^{\mathbf{q}v} + 1, r_{\tau_{\mathbf{p}v}, l}^{\mathbf{q}v}, l \in \mathcal{I}_v \setminus \{\rho\} \right)}{f_{\tau_{\mathbf{p}v}}^{\mathbf{p}v} \left(r_{\tau_{\mathbf{p}v}, \sigma}^{\mathbf{q}v} + 1, r_{\tau_{\mathbf{p}v}, l}^{\mathbf{q}v}, l \in \mathcal{I}_v \setminus \{\sigma\} \right)} \\
 & \prod_{w \in \mathbf{c}'(v)} \frac{f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho)}^{\mathbf{p}w} \left(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho), \tau_w}^{\mathbf{q}w} + 1, r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho), l}^{\mathbf{q}w}, l \in \mathcal{I}_w \setminus \{\tau_w\} \right)}{f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma)}^{\mathbf{p}w} \left(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma), \tau_w}^{\mathbf{q}w} + 1, r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma), l}^{\mathbf{q}w}, l \in \mathcal{I}_w \setminus \{\tau_w\} \right)} \frac{f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma)}^{\mathbf{p}w} \left(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma), l}^{\mathbf{q}w}, l \in \mathcal{I}_w \right)}{f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho)}^{\mathbf{p}w} \left(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho), l}^{\mathbf{q}w}, l \in \mathcal{I}_w \right)} \\
 (54) \quad & = \frac{f_{\tau_{\mathbf{p}'v}}^{\mathbf{p}'v} \left(r_{\tau_{\mathbf{p}'v}, \rho}^{\mathbf{q}'v} + 1, r_{\tau_{\mathbf{p}'v}, l}^{\mathbf{q}'v}, l \in \mathcal{I}_v \setminus \{\rho\} \right)}{f_{\tau_{\mathbf{p}'v}}^{\mathbf{p}'v} \left(r_{\tau_{\mathbf{p}'v}, \sigma}^{\mathbf{q}'v} + 1, r_{\tau_{\mathbf{p}'v}, l}^{\mathbf{q}'v}, l \in \mathcal{I}_v \setminus \{\sigma\} \right)} \\
 & \prod_{w \in \mathbf{c}'(v)} \frac{f_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \rho)}^{\mathbf{p}'w} \left(r_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \rho), \tau_w}^{\mathbf{q}'w} + 1, r_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \rho), l}^{\mathbf{q}'w}, l \in \mathcal{I}_w \setminus \{\tau_w\} \right)}{f_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \sigma)}^{\mathbf{p}'w} \left(r_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \sigma), \tau_w}^{\mathbf{q}'w} + 1, r_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \sigma), l}^{\mathbf{q}'w}, l \in \mathcal{I}_w \setminus \{\tau_w\} \right)} \frac{f_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \sigma)}^{\mathbf{p}'w} \left(r_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \sigma), l}^{\mathbf{q}'w}, l \in \mathcal{I}_w \right)}{f_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \rho)}^{\mathbf{p}'w} \left(r_{(\tau_{\mathbf{p}'w} \setminus \{v\}, \rho), l}^{\mathbf{q}'w}, l \in \mathcal{I}_w \right)}
 \end{aligned}$$

(b) We now simplify (54) further by writing it for properly chosen sparse \underline{r} 's.

This will lead us to functional equations for functions defined on \mathcal{I}_v .

We define

$$\mathfrak{d}_v = \mathbf{p}'_v \cap \mathbf{c}_v \quad \text{and} \quad \mathfrak{d}'_v = \mathbf{p}_v \cap \mathbf{c}'_v.$$

Note that due to the separation property (37) at least one of them is not empty.

Without loss of generality let us assume that $\mathfrak{d}_v \neq \emptyset$. Fix $\xi_{\mathfrak{d}_v} \in \mathcal{I}_{\mathfrak{d}_v}$ such that $\xi_i \neq \tau_i$ for any $i \in \mathfrak{d}_v$. For any $l \in \mathcal{I}_v$ denote by $\underline{i}(l)$ the cell with labels

$$i_v = l, \quad i_{\mathfrak{d}_v} = \xi_{\mathfrak{d}_v}, \quad i_y = \tau_y \text{ for } y \notin \mathfrak{d}_v \cup \{v\}.$$

Define

$$x_l = r_{\underline{i}(l)} \quad l \in \mathcal{I}_v.$$

Consider any $\underline{r} = (r_{\underline{i}})$ such that $r_{\underline{i}} = 0$ for all $\underline{i} \notin \{\underline{i}(l), l \in \mathcal{I}_v\}$. Since $\mathbf{p}_v \cap \mathfrak{d}_v = \emptyset$, by (36)

$$(55) \quad r_{\tau_{\mathbf{p}v}, l}^{\mathbf{q}v} = x_l, \quad l \in \mathcal{I}_v.$$

Again, by (36) and since $\mathbf{p}'_v \supset \mathfrak{d}_v \neq \emptyset$ we have

$$(56) \quad r_{\tau_{\mathbf{p}'v}, l}^{\mathbf{q}'v} = 0, \quad l \in \mathcal{I}_v.$$

Moreover, for $l \in \mathcal{I}_w$ and $k \in \mathcal{I}_v$ (particularly for $k = \rho$ or $k = \sigma$, which we shall use here)

$$r_{(\tau_{\mathbf{p}w} \setminus \{v\}, k), l}^{\mathbf{q}w} = \begin{cases} x_k, & \text{if } \mathbf{p}_w \cap \mathfrak{d}_v = \emptyset \text{ and either } (w \notin \mathfrak{d}_v, \text{ and } l = \tau_w) \text{ or } (w \in \mathfrak{d}_v \text{ and } l = \xi_w), \\ 0, & \text{otherwise} \end{cases}$$

and

$$r_{(\tau_{\mathbf{p}'w} \setminus \{v\}, k), l}^{\mathbf{q}'w} = \begin{cases} x_k, & \text{if } \mathbf{p}'_w \cap \mathfrak{d}_v = \emptyset \text{ and either } (w \notin \mathfrak{d}_v, \text{ and } l = \tau_w) \text{ or } (w \in \mathfrak{d}_v \text{ and } l = \xi_w), \\ 0, & \text{otherwise.} \end{cases}$$

These last two observations imply that the products $\prod_{w \in \mathbf{c}_v}$ and $\prod_{w \in \mathbf{c}'_v}$ in (54) factor into a function of x_ρ and a function of x_σ . Therefore their quotient can be written as $a_{v, \rho}(x_\rho)/a_{v, \sigma}(x_\sigma)$. Note, that potentially these functions may depend of \mathbf{p} and \mathbf{p}' , but it will not impact our final result.

Moreover, by (55) and (56) it follows that (54) assumes the form

$$\frac{f_{\tau_{\mathbf{p}v}}^{\mathbf{p}v} (x_\rho + 1, x_l, l \in \mathcal{I}_v \setminus \{\rho\})}{f_{\tau_{\mathbf{p}v}}^{\mathbf{p}v} (x_\sigma + 1, x_l, l \in \mathcal{I}_v \setminus \{\sigma\})} = K_v \frac{a_{v, \rho}(x_\rho)}{a_{v, \sigma}(x_\sigma)},$$

where

$$K_v = \frac{f_{\tau_{\mathbf{p}'_v}}^{\mathbf{p}'_v}(1_\rho, 0_l, l \in \mathcal{I}_v \setminus \{\rho\})}{f_{\tau_{\mathbf{p}'_v}}^{\mathbf{p}'_v}(1_\sigma, 0_l, l \in \mathcal{I}_v \setminus \{\sigma\})}.$$

Since $\tau_{\mathbf{p}_v}$ was arbitrary in $\mathcal{I}_{\mathbf{p}_v}$ we conclude from Lemma 8.1 that for any $\underline{k} \in \mathcal{I}_{\mathbf{p}_v}$ either

$$(57) \quad f_{\underline{k}}^{\mathbf{p}_v}(z_l, l \in \mathcal{I}_v) = \frac{\prod_{l \in \mathcal{I}_v} \left(A_{l|\underline{k}}^{v|\mathbf{p}_v} \right)^{z_l}}{\left(\tilde{A}_{\underline{k}}^{\mathbf{p}_v} \right)^{|z|}},$$

where $\tilde{A}_{\underline{k}}^{\mathbf{p}_v} = \sum_{l \in \mathcal{I}_v} A_{l|\underline{k}}^{v|\mathbf{p}_v}$ (recall that $(A)^z = A(A+1) \dots (A+z-1)$ is the ascending Pochhammer symbol) or it is a product of univariate power functions

$$(58) \quad f_{\underline{k}}^{\mathbf{p}_v}(z_l, l \in \mathcal{I}_v) = \prod_{l \in \mathcal{I}_v} \left[A_{l|\underline{k}}^{v|\mathbf{p}_v} \right]^{z_l}.$$

However the latter case is impossible due to the parameter independence assumption which requires that the distribution of the random vector $(\mathbb{P}_{\mathbf{p}}(X_v = l | \mathbf{X}_{\mathbf{p}_v} = \underline{k}), l \in \mathcal{I}_v)$ is non-degenerate.

We now want to identify the functions $f_{\underline{k}}^{\mathbf{p}'_v}$, $\underline{k} \in \mathcal{I}_{\mathbf{p}'_v}$. If $\mathfrak{d}'_v \neq \emptyset$ we can repeat the argument used to derive $f_{\underline{k}}^{\mathbf{p}_v}$ and obtain an analogue of (57) with \mathbf{p} replaced by \mathbf{p}' . If $\mathfrak{d}'_v = \emptyset$ we need to do some more work. We will use another sparse \underline{r} with new $\underline{i}(l)$'s defined by substituting \mathfrak{d}'_v for \mathfrak{d}_v . Note that under this new sparsity pattern for any $k \in \mathcal{I}_v$ (particularly for $k = \rho$ or $k = \sigma$, which we shall use here)

$$r_{(\tau_{\mathbf{p}_w} \setminus \{v\}, k), l}^{\mathbf{q}_w} = r_{(\tau_{\mathbf{p}'_w} \setminus \{v\}, k), l}^{\mathbf{q}'_w} = \begin{cases} x_k & \text{if } l = \tau_w, \\ 0 & \text{if } l \neq \tau_w, \end{cases}$$

and

$$r_{\tau_{\mathbf{p}_v}, l}^{\mathbf{q}_v} = r_{\tau_{\mathbf{p}'_v}, l}^{\mathbf{q}'_v} = x_l, \quad l \in \mathcal{I}_v.$$

Thus, (54) becomes

$$\frac{f_{\tau_{\mathbf{p}_v}}^{\mathbf{p}_v}(x_\rho + 1, x_l, l \in \mathcal{I}_v \setminus \{\rho\})}{f_{\tau_{\mathbf{p}_v}}^{\mathbf{p}_v}(x_\sigma + 1, x_l, l \in \mathcal{I}_v \setminus \{\sigma\})} = \frac{a_{v,\rho}(x_\rho)}{a_{v,\sigma}(x_\sigma)} \frac{f_{\tau_{\mathbf{p}'_v}}^{\mathbf{p}'_v}(x_\rho + 1, x_l, l \in \mathcal{I}_v \setminus \{\rho\})}{f_{\tau_{\mathbf{p}'_v}}^{\mathbf{p}'_v}(x_\sigma + 1, x_l, l \in \mathcal{I}_v \setminus \{\sigma\})}.$$

Plugging (57) into the left hand side above we obtain

$$\frac{f_{\tau_{\mathbf{p}'_v}}^{\mathbf{p}'_v}(x_\rho + 1, x_l, l \in \mathcal{I}_v \setminus \{\rho\})}{f_{\tau_{\mathbf{p}'_v}}^{\mathbf{p}'_v}(x_\sigma + 1, x_l, l \in \mathcal{I}_v \setminus \{\sigma\})} = \frac{A_{\rho|\tau_{\mathbf{p}_v}}^{v|\mathbf{p}_v} + x_\rho}{A_{\sigma|\tau_{\mathbf{p}_v}}^{v|\mathbf{p}_v} + x_\sigma} \frac{a'_{v,\sigma}(x_\sigma)}{a'_{v,\rho}(x_\rho)} \text{ or } \frac{A_{\rho|\tau_{\mathbf{p}_v}}^{v|\mathbf{p}_v}}{A_{\sigma|\tau_{\mathbf{p}_v}}^{v|\mathbf{p}_v}} \frac{a'_{v,\sigma}(x_\sigma)}{a'_{v,\rho}(x_\rho)},$$

respectively. Again we use Lemma 8.1 to conclude that one of the representations (57) or (58) (with \mathbf{p} changed into \mathbf{p}') holds also for $f_{\underline{k}}^{\mathbf{p}'_v}$ for any $\underline{k} \in \mathcal{I}_{\mathbf{p}'_v}$. Similarly, as above, we conclude that non-degeneracy implies that the representation given in (58) is not a valid one.

Given an arbitrary $v \in V$, so far, we have derived the expression of $f_{\underline{k}}^{\mathbf{p}_v}$ and $f_{\underline{k}}^{\mathbf{p}'_v}$ in (57) for an arbitrary separating pair $\mathbf{p}, \mathbf{p}' \in \mathcal{P}$. Clearly, (57) is valid for any $\mathbf{p} \in \mathcal{P}$ and any $v \in V$. Indeed, given $v \in V$ and the separating pair \mathbf{p}, \mathbf{p}' , consider another \mathbf{p}'' . Then, either $\mathbf{p}''(v) \neq \mathbf{p}(v)$ and then we consider the pair \mathbf{p}, \mathbf{p}'' or $\mathbf{p}''(v) = \mathbf{p}(v)$ and then we consider the pair $\mathbf{p}', \mathbf{p}''$.

Now, returning to (50) we see that for any d -way table $\underline{r} = (r_{\underline{i}}, \underline{i} \in \mathcal{I})$ of non-negative numbers, any $\mathbf{p} \in \mathcal{P}$, any $v \in V$ and any $(\underline{k}, l) \in \mathcal{I}_{q_v}$ there exist numbers $A_{l|\underline{k}}^{v|\mathbf{p}_v}$ such that

$$(59) \quad \mathbb{E} \prod_{\underline{i} \in \mathcal{I}} [\mathbb{P}_{\mathbf{p}}(\mathbf{X} = \underline{i})]^{r_{\underline{i}}} = \prod_{v \in V} \prod_{\underline{k} \in \mathcal{I}_{p_v}} \frac{\prod_{l \in \mathcal{I}_v} (A_{l|\underline{k}}^{v|\mathbf{p}_v})^{r_{\underline{k}, l}^{\mathbf{p}_v, v}}}{(\tilde{A}_{\underline{k}}^{\mathbf{p}_v})^{r_{\underline{k}}^{\mathbf{p}_v}}},$$

where

$$(60) \quad \tilde{A}_{\underline{k}}^{\mathbf{p}_v} = \sum_{l \in \mathcal{I}_v} A_{l|\underline{k}}^{v|\mathbf{p}_v}.$$

□

8.3. An auxiliary result on a functional equation.

Lemma 8.1. *Let F be a positive function defined on n th cartesian product of non-negative integers such that $F(\underline{0}) = 1$ and*

$$(61) \quad F(\underline{x}) = \sum_{i=1}^n F(\underline{x} + \underline{\epsilon}_i),$$

where $\underline{\epsilon}_i$ has all components equal to 0 except i th component which is 1. Assume that for any distinct $p, q \in \{1, \dots, n\}$

$$(62) \quad \frac{F(\underline{x} + \underline{\epsilon}_p)}{F(\underline{x} + \underline{\epsilon}_q)} = \frac{h_p(x_p)}{h_q(x_q)} \quad \forall \underline{x} = (x_1, \dots, x_n) \in \{0, 1, \dots\}^n$$

for some functions h_i , $i = 1, 2, \dots, n$.

Then there exists a vector $\underline{A} = (A_1, \dots, A_n) \in \mathbb{R}^n$ such that $\forall \underline{x} = (x_1, \dots, x_n) \in \{0, 1, \dots\}^n$ either

$$F(\underline{x}) = \frac{\prod_{i=1}^n (A_i)^{x_i}}{(|\underline{A}|)^{|\underline{x}|}},$$

where $|\underline{u}| = u_1 + \dots + u_n$ for any vector $\underline{u} = (u_1, \dots, u_n)$ and $(a)^m = a(a+1) \dots (a+m-1)$, or

$$F(\underline{x}) = \prod_{i=1}^n x_i^{A_i}.$$

Lemma 8.1, as given above, is a special version of Lemma 3.1 from Sakowicz and Wołowski (2014) (it suffices just to take $A + \{1, \dots, n\}$ in this lemma). It is also closely related to the argument used in the proof of Theorem 2 in Bobecka and Wołowski (2009).

8.4. Proof of (54).

1. Let $A_1(\rho)$ and $A_1(\sigma)$ be the values of $A_1 = \prod_{\underline{k} \in \mathcal{I}_{p_v} \setminus \{\tau_v\}} f_{\underline{k}}^{\mathbf{p}_v}(r_{\underline{k}, l}^{\mathbf{q}_v}, l \in \mathcal{I}_v)$, the second factor in the first line of the left hand side of (52), for $\tau_v = \rho$ and $\tau_v = \sigma$, respectively. Let $A'_1(\rho)$ and $A'_1(\sigma)$ be the analog quantities for the right-hand side of (52). Clearly

$$A_1(\rho) = \prod_{\underline{k} \in \mathcal{I}_{p_v} \setminus \{\tau_{p_v}\}} f_{\underline{k}}^{\mathbf{p}_v}(r_{\underline{k}, \rho}^{\mathbf{q}_v}, r_{\underline{k}, \sigma}^{\mathbf{q}_v}, r_{\underline{k}, l}^{\mathbf{q}_v}, l \in \mathcal{I}_v \setminus \{\rho, \sigma\}) = A_1(\sigma).$$

Similarly $A'_1(\rho) = A'_1(\sigma)$.

2. Let $A_2(\rho)$ and $A_2(\sigma)$ be the values of the factor for fixed $w \in \mathbf{c}_v$ in the second line of the left hand side of (52), for $\tau_v = \rho$ and $\tau_v = \sigma$, respectively. Let $A'_2(\rho)$ and $A'_2(\sigma)$ be the analog quantities for the right-hand side of (52). Clearly

$$A_2(\rho) = f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho)}^{\mathbf{p}w}(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho), \tau_w}^{\mathbf{q}w} + 1, r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho), l}^{\mathbf{q}w}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \\ f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma)}^{\mathbf{p}w}(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma), l}^{\mathbf{q}w}, l \in \mathcal{I}_w) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}w} \setminus \{\tau_{\mathbf{p}w}, (\tau_{\mathbf{p}w} \setminus \{v\}, \sigma)\}} f_{\underline{k}}^{\mathbf{p}w}(r_{\underline{k}, l}^{\mathbf{q}w}, l \in \mathcal{I}_w)$$

and

$$A_2(\sigma) = f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma)}^{\mathbf{p}w}(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma), \tau_w}^{\mathbf{q}w} + 1, r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \sigma), l}^{\mathbf{q}w}, l \in \mathcal{I}_w \setminus \{\tau_w\}) \\ f_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho)}^{\mathbf{p}w}(r_{(\tau_{\mathbf{p}w} \setminus \{v\}, \rho), l}^{\mathbf{q}w}, l \in \mathcal{I}_w) \prod_{\underline{k} \in \mathcal{I}_{\mathbf{p}w} \setminus \{\tau_{\mathbf{p}w}, (\tau_{\mathbf{p}w} \setminus \{v\}, \rho)\}} f_{\underline{k}}^{\mathbf{p}w}(r_{\underline{k}, l}^{\mathbf{q}w}, l \in \mathcal{I}_w).$$

Note that the two sets appearing in the indices: $\{\tau_{\mathbf{p}w}, (\tau_{\mathbf{p}w} \setminus \{v\}, \sigma)\}$ in $A_2(\rho)$ and $\{\tau_{\mathbf{p}w}, (\tau_{\mathbf{p}w} \setminus \{v\}, \rho)\}$ in $A_2(\sigma)$ are identical. Therefore, for each $w \in \mathbf{c}_v$ the ratio $\frac{A_2(\rho)}{A_2(\sigma)}$ is equal to the factor in the second line of the left hand side of (54). Similarly, for each $w \in \mathbf{c}_v$ the ratio $\frac{A'_2(\rho)}{A'_2(\sigma)}$ is equal to the factor in the second line of the right-hand side of (54).

3. Since v appears in the third line of neither the left hand side nor the right-hand side of (52), these lines cancel out in the ratios of (53).

DEPARTMENT OF MATHEMATICS AND STATISTICS, YORK UNIVERSITY, TORONTO, M3J1P3, CANADA
E-mail address: massamh@yorku.ca

WYDZIAŁ MATEMATYKI I NAUK INFORMACYJNYCH, POLITECHNIKA WARSZAWSKA, WARSZAWA, 00662, POLAND
E-mail address: wesolo@mini.pw.edu.pl